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*. Under compensatory rules a consumer will evaluate an item over all related aspects, thus a good performance on one aspect of an item compensates for poor performances on other aspects. For example, Alice wants to buy a cellphone and she is concerned about three relevant aspects: battery life, price and storage space. A compensatory rule to evaluate a cellphone is to score its performance on each aspect separately and compute a weighted summation over all aspects.

However, in the study of human choice behavior, it is well regarded that consumers more frequently make consumption related choices based on *non-compensatory rules* (Engel, Blackwell, and Miniard 1986). For example, (Hauser, Ding, and Gaskin 2009) reviews 132 empirical surveys in literature and concludes that more than 70% of consumers adopt non-compensatory rules when buying air-conditioners, automobiles, computers, cameras and so on.

Non-compensatory rules do not allow the shortcomings of a product to be balanced out by its attractive features. The literature has proposed different non-compensatory rules, among which *lexicographic rule* and *conjunctive rule* are the most common. For example, in a survey interviewing consumption decisions about beer brands and fast-food outlets (Laroche, Kim, and Matsui 2003), conjunctive rule has a success rate of 62.0% in predicting brand consideration and lexicographic rule has a success rate of 34.6% which is the second highest non-compensatory rule. We next illustrate *lexicographic rule* and *conjunctive rule* by the cellphone example in Table. 1.

Example. *Lexicographic rule* assumes that aspects of products can be ordered in terms of importance and alternative brands are evaluated sequentially from most prominent to least prominent aspects. If Alice's priority is long-lasting battery, then she will adopt lexicographic rule, to rank phones first based on 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 . *Conjunctive rule* establishes a minimally acceptable threshold for each aspect and evaluation is made on the basis of whether or not the products satisfy the threshold. If Alice wants the phone to be cheap and with plenty of storage space, then she will adopt conjunctive rule, to set thresholds (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 is problematic.

Table 1: The cellphone example to illustrate non-compensatory rules.

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

Current computer support for non-compensatory rules is to manually control the rules in decision support systems (Lee 2009) , e.g. consumers are asked to specify the threshold for an aspect of interest. Such manual approaches are labor costly and difficult to set up and maintain. On the contrary, learning models such as latent factor models (Koren, Bell, and Volinsky 2009) have the advantage of scalability, simplicity and flexibility. Unfortunately, to the best of our knowledge, no previous effort has been devoted to building learning to recommend models based on non-compensatory rules.

Our goal in this paper is to study this unexplored area of how learning models can be benefited by non-compensatory rules. Two challenges needed to be addressed. (1) How to embed symbolic non-compensatory rules in a data-driven learning framework? (2) How to present a general framework based on non-compensatory rules in a manner that complies with a variety of existing recommendation models, including rating prediction and ranking aware models?

Our primary contribution is to give a conceptual model of how users adopt non-compensatory rules in recommender systems. Our assumptions are based on the lexicographic and conjunction rules. 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 BT 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. We start with surveying the most commonly adopted latent factor models in the community of recommendation research. We show that previous research work are compensatory models which are based on different rating prediction formulas and ranking models. Next, we describe our non-compensatory assumptions and develop non-compensatory versions of existing models. Then we experimentally show that the non-compensatory versions outperform the original versions of existing models on a variety of real-world data sets. Finally we conclude our work and future directions.

Compensatory Recommendation Models

In this section, we show that existing latent factor models are based on compensatory rules. We summarize and categorize related work based on the forms of 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, which are rows of item space $\mathbf{V} \in \mathcal{R}^{N\times K}$, $\mathbf{u} \in \mathcal{R}^K$ denotes the user preferences, which is a row of the user space $\mathbf{U} \in \mathcal{R}^{M\times K}$. $\mathbf{U}, \mathbf{V}$ are components of the model parameters $\Theta = \{\mathbf{U}, \mathbf{V}\}$.

Rating Prediction Formulas

One goal of recommendation research is to recover the rating matrix $\mathbf{X}$, by minimizing a loss function $\mathcal{L}(\Theta)$, 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.

$$\mathcal{L}(\Theta) = \sum_{u,q} (\mathbf{X}_{u,q} - \hat{\mathbf{X}}_{u,q})^2 + \lambda(\|\mathbf{U}\| + \|\mathbf{V}\|) \quad (1)$$

We list some of the most successful rating prediction formulas for $\hat{\mathbf{X}}$.

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

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \mathbf{q}_k \mathbf{u}_k \quad (2)$$

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. 2. 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.

Table 2: Existing latent factor models in literature can be classified based on the loss functions and rating prediction formulas

Loss Function	Rating Prediction Formula		
	Matrix Factorization	Neighborhood Factorization	Local Low-rank Factorization
Point-wise Square Loss	MF (Koren, Bell, and Volinsky 2009)	AMF (Koren 2008)	LLORMA (Lee et al. 2013)
Pair-wise Thurstone Ranking	BPR (Rendle et al. 2009)	FSBPR (Zhao et al. 2018)	LCR (Lee et al. 2014)
Pair-wise BT Ranking	BT (Hu and Li 2016)		

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 Asymmetric Matrix Factorization (AMF) in (Koren 2008) is stated as follows.

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \mathbf{q}_k \Big(\sum_{p \in R(u)} \mathbf{p}_k / \sqrt{|R(u)|} \Big), \tag{3}$$

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 Local Low-Rank Matrix Approximation (LLORMA) (Lee et al. 2013). The intuition 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:

$$\hat{\mathbf{X}}_{u,q} = \sum_{t=1}^{S} \sum_{k} \mathbf{u}_{t,k} \frac{K((\mathbf{u}_t, \mathbf{i}_t), (\mathbf{u}, \mathbf{q}))}{\sum_{s=1}^{S} K((\mathbf{u}_s, \mathbf{i}_s), (\mathbf{u}, \mathbf{q}))} \mathbf{q}_{t,k} \tag{4}$$

$\mathbf{u}_t, \mathbf{q}_t$ 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 $\mathcal{L}(\Theta)$ on the observed pair-wise rankings. For example, the Bayesian posterior is expressed as:

$$\mathcal{L}(\Theta) = \sum_{u} \sum_{p,q} o(p \succ_u q) \log p(p \succ_u q) + \lambda(\|\mathbf{U}\| + \|\mathbf{V}\|). \tag{5}$$

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.

$$p(p \succ_u q) = \frac{1}{1 + \exp[-(\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q})]} \tag{6}$$

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.

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

$$p(p \succ_u q) = \frac{\exp \hat{\mathbf{X}}_{u,p}}{\exp \hat{\mathbf{X}}_{u,p} + \exp \hat{\mathbf{X}}_{u,q}} \tag{7}$$

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 BT 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 the next section 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. We proceed to present a general framework for modeling the psychological assumptions about decision rules. We show the universality of the proposed framework by realizing it in different rating prediction formulas and ranking models.

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, while in non-compensatory rules factors are not interchangeable. (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.

For computational convenience, inspired by the psychological findings, we present the following conceptual model based on lexicographic and conjunction rules. 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. Two types of evaluation strategies are adopted, one for the prominent aspect and the other for other non-prominent aspects. The overall evaluation of the item is mainly based on the its performance on the prominent aspect. The overall evaluation is less influenced by the item's performance on non-prominent aspects, compared with the user-defined aspect-specific threshold.

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 threshold on aspect k set by user u is denoted as $\mathbf{b}_{u,k}$. When the aspect k is chosen, the evaluation of user u on q is $\exp\theta\mathbf{q}_k + \sum_{k'\neq k}(\mathbf{q}_{k'} - \mathbf{b}_{u,k'})$. 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-N

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}} [\exp\theta\mathbf{q}_k + \sum_{k'\neq k}(\mathbf{q}_{k'} - \mathbf{b}_{u,k'})]. \tag{8}$$

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

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \frac{\exp(\sum_{p\in R(u)} \mathbf{p}_k)}{\sum_{k'} \exp(\sum_{p\in R(u)} \mathbf{p}_{k'})} [\exp\theta\mathbf{q}_k + \sum_{k'\neq k}(\mathbf{q}_{k'} - \mathbf{b}_{u,k'})]. \tag{9}$$

LLORMA-N uses the same decomposition for each submatrix.

$$\hat{\mathbf{X}}_{u,q} = \sum_{t=1}^{S} \sum_k \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}} \frac{K((\mathbf{u}_t,\mathbf{i}_t),(\mathbf{u},\mathbf{q}))}{\sum_{s=1}^{S} K((\mathbf{u}_s,\mathbf{i}_s),(\mathbf{u},\mathbf{q}))} [\exp\theta\mathbf{q}_{t,k} + \sum_{k'\neq k}(\mathbf{q}_{t,k'} - \mathbf{b}_{u,k'})] \tag{10}$$

We can see that all these N versions are combinations of lexicographic and conjunction rules, where $\exp\theta \to \infty$ indicates that the user adopts lexicographical rules only. The threashold for a user on an aspect is static in the sense that $\mathbf{b}_{u,k}$ does not change by the nature of the items.

Non-Compensatory Ranking Models

Thurston-N. The modification of Thurston model is straightforward, as the ranking probability involves a subtraction component of $\hat{\mathbf{X}}_{u,q}$ which can be replaced by any N-version of rating prediction formulas. Note that the user-defined aspect specific threshold $\mathbf{b}_{u,k}$ cancels between $\hat{\mathbf{X}}_{u,p}$ and $\hat{\mathbf{X}}_{u,q}$.

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:

$$\mathcal{L} = -\sum_u \sum_{p\succ_u q} \ln \frac{1}{1+\exp -[\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}]} - \lambda\|\Theta\|, \tag{11}$$

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 $\Delta\hat{\mathbf{X}}_{u,p,q} = \hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}$.

BT-N. Finally we propose the non-compensatory version of BT 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 BT 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.

$$p(p \succ_u q) = \prod_{k=1}^{K} \mathbf{u}_k [\frac{\mathbf{p}_k}{\mathbf{p}_k + \theta\mathbf{q}_k} \prod_{k'\neq k} \frac{\theta\mathbf{p}_{k'}}{\mathbf{q}_{k'} + \theta\mathbf{p}_{k'}}]. \tag{12}$$

where $\mathbf{u}_k > 0, \sum_k \mathbf{u}_k = 1, \mathbf{p}, \mathbf{q} > \mathbf{0}$ and $\theta > 1$. BT-N 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 $\forall k' \neq k, \theta p_{k'} > q_{k'}, \theta > 1$. BT-N 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.

[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.

To infer the parameters of BT-N, 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

$$k \sim u_k^t \frac{\mathbf{p}_k^t}{\mathbf{p}_k^t + \theta^t \mathbf{q}_k^t} \prod_{k' \neq k} [\frac{\theta^t \mathbf{p}_{k'}^t}{\mathbf{q}_{k'}^t + \theta^t \mathbf{p}_{k'}^t}]. \tag{13}$$

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 N versions of rating prediction models outperform the original versions on rating data sets. The second set of experiments is conducted to examine whether N versions of ranking aware models outperform the original versions on data sets with explicit rating feedback. The third set of experiments is conducted to examine whether N versions of ranking aware models outperform the original versions on data sets with graded implicit feedback. We also analyze the inferred parameters θ, $\mathbf{b}$ in non-compensatory rules for further insights.

Comparative Results for Rating Prediction Models

Data Sets We use the standard benchmarking datasets with user-item ratings. (1) Movielens[2]: user-movie rating for movies collected from the MovieLens web site (2) FilmTrust (Guo, Zhang, and Yorke-Smith 2013): user-movie ratings crawled from the entire FilmTrust website. (3) CiaoDVD (Guo et al. 2014): user-movie ratings crawled from the entire category of DVDs from the UK Ciao website. Statistics of the datasets are described in Table. 3.

For each dataset, we reserve users with at least 5 ratings and randomly split 80% of the ratings as training and 20% as test set. We avoid cold-start users and items. We consider each rating as an individual evaluation session. The ratings are normalized to the range of $[0, 1]$. The reported results are averaged using 5-fold cross validation,

Table 3: Statistics of Datasets with ratings

Dataset	#users	#items	#ratings	#pairs
Movielens	942	1,650	80,000	4,641,262
FilmTrust	1,235	2,062	35,497	623,516
CiaoDVD	2,665	14,280	72,665	2,478,836

Comparative Methods. We compare the non-compensatory improved versions (with suffix "-N") with the original versions on three widely adopted rating prediction methods (1) MF (Koren, Bell, and Volinsky 2009): standard matrix factorization in Equ. 2; (2) AMF (Koren 2008): neighborhood factorization in Equ. 3; (3) LLORMA (Lee et al. 2013): local low-rank matrix factorization in Equ. 4. For all methods including the N versions, we set the number of aspects $K = 10$. The regularization coefficients for MF and MF-N is $\lambda = 0.01$. The number of local models for LLORMA and LLORMA-N is $S = 10$. The learning rate is self adapted as in (Wilson and Martinez 2003). We stop the learning process either when the improvement in training error is smaller than $1e^{-6}$ or when the algorithm reaches 1000 iterations. To reduce the number of parameters, we set the user-defined aspect-specific threshold for the N models $\mathbf{b}_{u,k} = 0$ for every u, k.

Evaluation Metrics. We evaluate different approaches based on the following metrics. (1) AUC: first computes the area under precision-recall curve based on the predicted ratings for each user, then averaged over all users; (2) NDCG: another evaluation metric to measure the accuracy of item ranking per user by the predicted ratings v.s. the actual ranking, averaged over all users; (3) MRR: computes the reciprocal of the position of the item with the largest observed rating in the predicted ranking for each user, averaged over all users.

Table 4: Comparative performance for compensatory and non-compensatory rating prediction models, 'Improve' indicates the improvements of non-compensatory versions relative to the original models.

Dataset	Method	AUC	Improve	NDCG	Improve	MRR	Improve
Movie lens	MF	0.6661		0.6856		0.8391	
	MF-N	0.6990	4.94%	0.7107	3.67%	0.8745	4.23%
	AMF	0.6043		0.5003		0.7506	
	AMF-N	0.6129	1.43%	0.5027	0.48%	0.7559	0.71%
	LLORMA	0.6453		0.8990		0.5761	
	LLORMA-N	0.6516	0.98%	0.8994	0.04%	0.5761	0.01%
Film trust	MF	0.6056		0.5240		0.7522	
	MF-N	0.6166	1.81%	0.5252	0.24%	0.7624	1.35%
	AMF	0.6244		0.5055		0.7622	
	AMF-N	0.6436	3.07%	0.5098	0.85%	0.7717	1.24%
	LLORMA	0.5336		0.8672		0.6481	
	LLORMA-N	0.5517	3.39%	0.8684	0.14%	0.6533	0.80%
Ciao DVD	MF	0.5915		0.6497		0.8427	
	MF-N	0.6948	17.45%	0.6872	5.77%	0.8969	6.43%
	AMF	0.6211		0.5048		0.7607	
	AMF-N	0.7993	28.69%	0.5657	12.05%	0.8950	17.67%
	LLORMA	0.6986		0.7827		0.4883	
	LLORMA-N	0.6998	0.17%	0.7838	0.14%	0.4904	0.43%

We can see from Table. 4 that overall adopting non-compensatory rules can improve model performance. We observe that for "simpler" models, i.e. MF and AMF, the improvement is more significant. For complicated models such as LLORMA, the improvement is less significant. The reason is that LLORMA approximates the observations by several low-rank factorizations in different local neighborhoods. Thus LLORMA implements several layers of compensatory rules. While compensatory rule in its nature is addictive and extendable in a layered computation, non-compensatory rules may not fit perfectly in the layered framework. However, increasing the model complexity also leads to increased computation time and poor interpretability. Thus utilizing non-compensatory rules in simpler models, such as MF and AMF, generates recommendations with higher accuracy, efficiency and interpretability.

[2] http://www.grouplens.org

Comparative Results for Ranking Models

Data Sets. Next we evaluate the performance of models that target to ranking reconstruction. The datasets used are again Movielens, Filmtrust and CiaoDVD. We construct pair-wise ordering for each user between any higher rated item and lower rated item, i.e. $\mathbf{X}_{u,p} > \mathbf{X}_{u,q} \to p \succ_u q$. The number of ranking pairs on each dataset is shown in Table. 3

Comparative Methods. We compare the N improved versions with the original versions on four widely adopted ranking methods. (1) BT (Hu and Li 2016);: the Bradley-Terry ranking model with MF rating prediction formula, (2) BPR (Rendle et al. 2009): the Thurstonian ranking model with MF rating prediction formula, the optimization is through maximal Bayesian posterior, the regularization coefficient is $\lambda = 0.01$, (3) FSBPR (Zhao et al. 2018): the Thurstonian ranking model with AMF rating prediction formula, the optimization is through maximal Bayesian posterior, (4) LCR (Lee et al. 2014): the Thurstonian ranking model with local low-rank matrix factorization, the loss function for LLORMA and LLORMA-N is $log[M]$ which is the log-likelihood. The number of local models is $S = 10$. For all methods including the N versions, we set the number of aspects $K = 10$.

Evaluation Metrics. The goal is to reconstruct the observed rankings for each user. Hence we adopt the same set of ranking evaluation metrics, including AUC, NDCG and MRR.

Table 5: Comparative performance for compensatory and non-compensatory ranking aware models, 'Improve' indicates the improvements of non-compensatory versions relative to the original models.

Dataset	Method	AUC	Improve	NDCG	Improve	MRR	Improve
Movie Lens	BT	0.6021		0.5070		0.7654	
	BT-N	0.6434	6.86%	0.5425	7.02%	0.8440	10.27%
	BPR	0.7002		0.5443		0.8478	
	BPR-N	0.7246	3.48%	0.5508	1.20%	0.8623	1.71%
	FSBPR	0.5923		0.4993		0.7484	
	FSBPR-N	0.6512	9.94%	0.5205	4.26%	0.7954	6.28%
	LCR	0.6187		0.8317		0.6793	
	LCR-N	0.6197	0.16%	0.8359	0.50%	0.6862	1.02%
Film Trust	BT	0.6317		0.5070		0.7654	
	BT-N	0.7269	15.07%	0.5312	4.79%	0.8190	7.00%
	BPR	0.7825		0.5147		0.7825	
	BPR-N	0.6728	-14.02%	0.5392	4.75%	0.8365	6.91%
	FSBPR	0.5970		0.4996		0.7490	
	FSBPR-N	0.7165	20.01%	0.5205	4.20%	0.7954	6.19%
	LCR	0.6931		0.9480		0.8122	
	LCR-N	0.7001	1.01%	0.9503	0.24%	0.8232	1.35%
Ciao DVD	BT	0.6136		0.5230		0.8008	
	BT-N	0.7892	28.62%	0.5857	11.99%	0.9393	17.30%
	BPR	0.6380		0.4883		0.7240	
	BPR-N	0.8982	40.78%	0.5922	21.28%	0.9537	31.71%
	FSBPR	0.6003		0.5001		0.7502	
	FSBPR-N	0.6589	9.76%	0.5637	12.72%	0.8908	18.73%
	LCR	0.5211		0.9044		0.7515	
	LCR-N	0.5252	0.79%	0.9050	0.07%	0.7527	0.16%

A general observation in Table. 5 is that embedding non-compensatory rules significantly improves existing ranking models. In terms of AUC, NDCG and MRR, the non-compensatory models outperform the compensatory models on all datasets, with the only exception of BPR-N in AUC on Filmtrust dataset. The results validate the adequacy of non-compensatory rules in ranking aware models. Furthermore, by comparing Table. 5 and Table. 4, we observe that the non-compensatory rules generally make bigger improvements on ranking aware models than on rating aware models. This observation indicates that it is possible that consumers adopt non-compensatory rules more often in ranking alternative products.

Ranking Performance for Graded Implicit Feedback

In most recommender systems, users not only give explicit ratings but also implicit feedback that can be graded. For example, a purchase and a click are both implicit feedback that indicates user preference. A reasonable grading is that a purchase is "higher" than a click, as a purchase is a stronger indicator of user preference. Therefore, we conduct experiments on datasets with graded implicit feedback.

Table 6: Statistics of Datasets with graded implicit feedback

Dataset	#users	#items	#pairs	#sessions
Tmall-single	33,815	176,231	5,682,833	364,844
Tmall-hybrid	62,101	198,344	6,072,061	475,503
Yoochoose	341,396	30,852	3,044,572	341,396

Data Sets We use three real world datasets, as shown in Table. 6. Tmall[3] is a collection of user shopping sessions, where in each session the user has four types of activities: click, add to cart, add to favorite and purchase. We build two data sets based on Tmall. (1) Tmall-single: a set of pairwise rankings where an item p purchased in u's session is considered to be superrior than an item q clicked in the same session. (2) Tmall-hybrid: the pairwise rankings are built by extracting purchased items in each session and all remaining items which are not purchased in the same session. Thus if an item p is purchased in the session, and an item q is either clicked, added to cart or added to favorite, we build $p \succ_u q$. (3) Yoochoose[4]: a collection of user shopping sessions with clicked and purchased items. In this data set, user information is not provided. In the experiments, we assume that each session is from a new user.

Comparative Methods. We compare the N improved versions with the original versions on the same four ranking models. It is worthy to note that implementation of BT-N is different from previous sections. In our model the prominent aspect is associate with each evaluation session. In the previous experiments, an evaluation session is a rating or a pair of rating. Here we the user interaction session information is available. Thus in BT-N, we sample the prominent aspect for each session instead of a pair of actions.

Evaluation Metrics. In addition to the aforementioned ranking evaluation metrics: AUC, NDCG and MRR, in order to evaluate the sessional ranking performance, we adopt two more evaluation metrics: MAP and Precision. MAP first computes the mean precision at each position of the predicted ranking per session, then averages it over all sessions. Precision first computes the fraction of correctly ordered test pairs in each session, then averages it over all sessions.

[3] https://ijcai-15.org/index.php/repeat-buyers-prediction-competition

[4] http://2015.recsyschallenge.com

Table 7: Comparative performance for compensatory and non-compensatory models on ranking implicit feedback, 'Improve' indicates the improvements of non-compensatory versions relative to the original models.

Dataset	Method	AUC	Improve	NDCG	Improve	MRR	Improve	MAP	Improve	Prec	Improve
Tmall-single	BT	0.7253		0.2814		0.4890		0.4348		0.2787	
	BT-N	0.7308	0.76%	0.2853	1.36%	0.4973	1.70%	0.4408	1.38%	0.2810	0.81%
	BPR	0.7252		0.2826		0.4932		0.4359		0.2789	
	BPR-N	0.7305	0.73%	0.2854	0.99%	0.4977	0.91%	0.4410	1.17%	0.2810	0.77%
	FSBPR	0.7092		0.2732		0.4717		0.4163		0.2734	
	FSBPR-N	0.7130	0.53%	0.2747	0.57%	0.4740	0.48%	0.4193	0.72%	0.2747	0.48%
	LCR										
	LCR-N										
Tmall-hybrid	BT	0.7931		0.3056		0.5458		0.5015		0.2934	
	BT-N	0.8249	4.01%	0.3305	8.15%	0.6063	11.09%	0.5592	11.50%	0.3044	3.75%
	BPR	0.8132		0.3248		0.5950		0.5463		0.3006	
	BPR-N	0.8267	1.66%	0.3324	2.33%	0.6112	2.72%	0.5635	3.15%	0.3050	1.45%
	FSBPR	0.7431		0.2770		0.4817		0.4398		0.2768	
	FSBPR-N	0.7624	2.60%	0.2865	3.44%	0.5007	3.95%	0.4597	4.53%	0.2831	2.24%
	LCR										
	LCR-N										
Yoochoose	BT	0.8732		0.4742		0.7156		0.6368		0.4569	
	BT-N	0.8966	2.67%	0.5166	8.94%	0.7882	10.14%	0.7112	11.69%	0.4786	4.76%
	BPR	0.8934		0.5019		0.7639		0.6821		0.4711	
	BPR-N	0.9034	1.12%	0.5149	2.59%	0.7854	2.82%	0.7057	3.46%	0.4777	1.39%
	FSBPR	0.8405		0.4379		0.6541		0.5685		0.4374	
	FSBPR-N	0.8839	5.15%	0.5445	24.34%	0.7987	22.11%	0.6825	20.06%	0.5362	22.59%
	LCR										
	LCR-N										

As shown in Table. 7, the non-compensatory models outperform the original models in terms of all evaluation metrics on all data sets. Thus it is safe to conclude that consumers also conduct non-compensatory rules in the process of giving implicit feedback. Comparing among all models, the improvement is particularly notable for BT model, which suggests that fixing the prominent aspect in a whole session is beneficial. Finally, we observe that FSBPR performs poorly on Yoochoose, possibly because FSBPR connects each user directly to its neighbors, thus FSBPR is not suitable for datasets with sparse user feedback. However, the improvement brought by non-compensatory rules to FSBPR is the most significant, showing its potential in handling cold-start users with a neighborhood factorization framework.

Analysis of Inferred Parameters

Finally, we analyze the values of inferred parameters $\theta, \mathbf{b}$ in N models to gain some insights about the non-compensatory rules. As the user-defined aspect-specific threshold $\mathbf{b}_{u,k}$ does not affect ranking aware models, we implement the non-compensatory matrix factorization (MF-N) model on three rating datasets, including Movielens, Filmtrust and CiaoDVD. We let $\theta, \mathbf{b}$ to be automatically learnt from the training data. We run 5-fold validations for 10 times, where each time we randomly splitting the data sets to 5 folds. We present in Table. 8 the mean and standard deviation of θ inferred on each dataset. We also present the mean and standard deviation of the standard deviation of $\mathbf{b}$ inferred on each dataset. That is, we first compute the standard deviation of $\sigma(\mathbf{b}_u) = \sqrt{[\sum_{k=1}^{K}(\mathbf{b}_{u,k} - \bar{\mathbf{b}}_u)^2/K]}$ for each user, where $\bar{\mathbf{b}}_u = \sum_{k=1}^{K} \mathbf{b}_{u,k}/K$ is the average threshold for user u over all aspects. Then we report the mean and standard deviation of σ_u over all users.

Strength of Non-compensatory Rules We have the following observations. (1) The obtained value $\theta > 0$ on all datasets. Since $\exp \theta > 1$, the prominent aspect is more important than non-prominent aspect in user evaluations. This is consistent to our assumptions that lexicographical rules will evaluate item performance first on the most important aspect. (2) The optimal value of θ is moderate, indicating the users adopt a combination of lexicographical rules and conjunctive rules.

Effect of User-defined Aspect-specific Threshold. We can see that the standard deviation is positive $\sigma(\mathbf{b}_u) > 0$, suggesting that the user defined threshold for each aspect is significant different. The experimental result is consistent with our assumption of employing aspect-specific threshold in non-compensatory rules.

Table 8: Scale of the strength of lexicographical rule θ and user-defined aspect-specific threshold $\mathbf{b}_{u,k}$.

Dataset	Movielens	FilmTrust	CiaoDVD
θ	0.608 ± 0.105	0.667 ± 0.016	0.773 ± 0.051
$\sigma(\mathbf{b}_u)$	0.131 ± 0.007	0.863 ± 0.072	0.058 ± 0.001

Conclusion

Psychology study has shown that consumers adopt compensatory and non-compensatory rules in the decision making process. However, all existing latent factor models in recommendation systems are essentially based on compensatory rules. In this contribution, we present for the first time in the literature of recommendation systems how non-compensatory rules can be embedded in latent factor models. We show that applying non-compensatory rules can universally boost recommendation performance for a variety of rating prediction and ranking aware models.

References

Devooght, R.; Kourtellis, N.; and Mantrach, A. 2015. Dynamic matrix factorization with priors on unknown values. In *Proceedings of the 21th ACM SIGKDD Interna-*

tional 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.

Guo, G.; Zhang, J.; Thalmann, D.; and Yorke-Smith, N. 2014. Etaf: An extended trust antecedents framework for trust prediction. In *Proceedings of the 2014 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 540–547.

Guo, G.; Zhang, J.; and Yorke-Smith, N. 2013. A novel bayesian similarity measure for recommender systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2619–2625.

Hauser, J. R.; Ding, M.; and Gaskin, S. P. 2009. Non-compensatory (and compensatory) models of consideration-set decisions. In *2009 Sawtooth Software Conference Proceedings, Sequin WA*.

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.

Laroche, M.; Kim, C.; and Matsui, T. 2003. Which decision heuristics are used in consideration set formation? *Journal of Consumer Marketing* 20(3):192–209.

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.

Wilson, D. R., and Martinez, T. R. 2003. The general inefficiency of batch training for gradient descent learning. *Neural Networks* 16(10):1429–1451.

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*.