Non-Compensatory Psychological Models for Recommender Systems

Chen Lin,[1,3] **Xiaolin Shen,**[1] **Si Chen**[1]**, Muhua Zhu**[3]**,** and **Yanghua Xiao**[2,3]

[1] Department of Computer Science, Xiamen University, Xiamen, Fujian, China
[2] School of Computer Science, Fudan University, Shanghai, China
[3] Alibaba Group, Hangzhou, Zhejiang, China
chenlin@xmu.edu.cn, {xlshen, sichen}@stu.xmu.edu.cn, muhua.zmh@alibaba-inc.com, shawyh@fudan.edu.cn

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. they evaluate an item over multiple aspects and compute a weighted or/and summated score which is used to further derive the rating or ranking of the item. However, it has been shown in the literature of consumer behavior that, consumers adopt non-compensatory rules more often than compensatory rules. Our main contribution in this paper is to study the unexplored area of utilizing 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 mainly based on item performance on the prominent aspect. For non-prominent aspects the user sets a minimal acceptable threshold. We give a conceptual model for these general assumptions. We show how this conceptual model can be realized in both point-wise rating prediction models and pair-wise ranking prediction models. Experiments on real-world data sets validate that adopting non-compensatory rules improves recommendation performance for both rating and ranking models.

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 products of the user preferences and the item features. There are fruitful 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 evaluates an item over all aspects, thus a good performance on one aspect of an item compensates for poor performances on other aspects. For example, there are concerns about three relevant aspects: battery life, price and storage space when Alice wants to buy a cellphone. 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 the 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 commonly used. 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 the most prominent to the 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. In this case 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. 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). In this case, iPhone fails to meet the cut-off point, thus it will not outrank Honor. In either case, adopting a compensatory rule

based recommendation model is problematic.

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

Item	Prominent aspect	Non-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 benefit from non-compensatory rules. Two challenges need to be addressed. (1) How to embed symbolic non-compensatory rules in a data-driven learning framework? (2) How to build 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 propose 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 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 local low-rank matrix approximation (LLORMA (Lee et al. 2013)), and pairwise ranking reconstruction models such as the Bradley-Terry model (BT (Hu and Li 2016)) and Thurstone model (BPR (Rendle et al. 2009)).

We conduct comprehensive experiments on a variety of real world data sets. We experimentally show that *the proposed non-compensatory framework universally improves recommendation performances of different existing models*.

The paper is organized as follows. We start with surveying the most commonly adopted latent factor models.

We show that previous research work is based on compensatory rules. Next, we present a conceptual model based on non-compensatory rules and realize it in different existing models. Then, we experimentally show that the non-compensatory modifications outperform the original versions of existing models on a variety of real-world data sets. Finally we conclude our work.

Compensatory Recommendation Models

In this section, we show that existing latent factor models can be categorized based on the forms of rating prediction formulas and loss functions. Table 2 summarizes typical related work. 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 rating aware 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 $\mathbf{X}_{u,q}$ for each user u who has rated item q.

$$\mathcal{L}(\Theta) = \sum_{u,q}(\mathbf{X}_{u,q} - \hat{\mathbf{X}}_{u,q})^2 + \lambda_U\|\mathbf{U}\|_2^2 + \lambda_V\|\mathbf{V}\|_2^2 \quad (1)$$

We list three 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.

Neighborhood Factorization. In traditional memory based collaborative filtering strategies, neighborhood information has been proved to be useful. It is possible to incorporate such neighborhood information in latent factor models. Instead of directly modeling user preferences $\mathbf{u}$, each

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
Square Loss	MF (Koren, Bell, and Volinsky 2009)	AMF (Koren 2008)	LLORMA (Lee et al. 2013)
Thurstone Ranking	BPR (Rendle et al. 2009)	FSBPR (Zhao et al. 2018)	LCR (Lee et al. 2014)
BT Ranking	BT (Hu and Li 2016)	-	-

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 $\mathbf{p}, \mathbf{q}$, where $\mathbf{q}$ is its own feature vector and $\mathbf{p}$ is an auxilary vector for each of its neighboring items. The preference vector of . The rating prediction formula of Asymmetric Matrix Factorization (AMF) in (Koren 2008) is:

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

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). Instead of assuming the entire rating matrix $\mathbf{X}$ is low-rank, the intuition of LLORMA is that only a sub-matrix is low-rank. The sub-matrix is restricted to a neighborhood. Therefore, the predicted rating is aggregated over S sub-matrices:

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

where $\mathbf{u}_t, \mathbf{q}_t$ 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 $p \succ_u q$, which indicates 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).

Thurstone Model The Thurstone model (Thurstone 1927) is widely used in recommendation systems. It uses a non-linear transformation of the predicted ratings $\hat{\mathbf{X}}$ to compute the probability of pair-wise rankings $Pr(p \succ_u q)$.

$$Pr(p \succ_u q) = \frac{1}{1 + \exp[-(\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q})]} \qquad (5)$$

Thurstone models in the community of recommendation systems aim to reconstruct Bayesian Personalized Ranking (BPR (Rendle et al. 2009)) by minimizing a negative Bayesian posterior function $\mathcal{L}(\Theta)$ on the observed pair-wise rankings.

$$\mathcal{L}(\Theta) = -\sum_u \sum_{p,q} I(p \succ_u q) \log Pr(p \succ_u q) \qquad (6)$$
$$+\lambda_U \|\mathbf{U}\|_2^2 + \lambda_V \|\mathbf{V}\|_2^2,$$

where $I(p \succ_u q)$ is an indicator function of whether or not the preference is observed.

Bradley-Terry Model. Another extensively studied model for learning to rank is Bradley-Terry model (Hunter 2004), which generates ranking pairs by a division:

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

where $\hat{\mathbf{X}}_{u,p}$ is the predicted rating by MF. The objective of BT model is the negative log-likelihood $\log Pr(p \succ_u q)$ on all observations:

$$\mathcal{L}(\Theta) = -\sum_u \sum_{p,q} I(p \succ_u q) \log Pr(p \succ_u q). \qquad (8)$$

There are other variants of loss functions. For example, penalty terms of the L2-norm of U, V, as well as the differences between $\hat{\mathbf{X}}$ and $\mathbf{X}$ are addded in an improved BT model (Hu and Li 2016).

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 into components which are directly based on each score $\hat{\mathbf{X}}_{u,q}$ 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. (3) As the key component $\hat{\mathbf{X}}_{u,q}$ is based on compensatory rules, all existing latent factor models are essentially compensatory.

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 non-compensatory 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 conceptual model in Fig. 1 based on lexicographic and conjunction rules. We assume that in each evaluation session (i.e. either a true user interaction session with multiple actions, or a pseudo session which contains one rating action), 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. Evaluation on the prominent aspect is enhanced by strength coefficient θ. Evaluations on non-prominent aspects are based on item performance, compared with the user-specific aspect-specific threshold $\mathbf{b}_u$. The overall evaluation, which could be either a rating or a ranking, aggregates over prominent and non-prominent aspects.
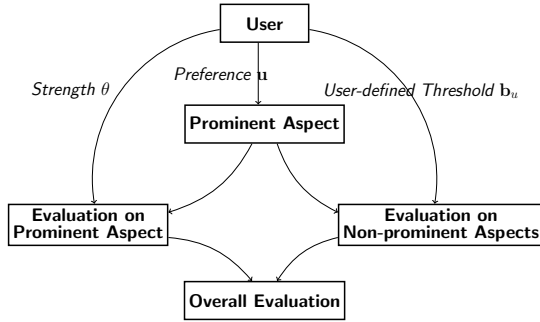

Figure 1: Conceptual model for the user evaluation process

Non-Compensatory Rating Prediction Formulas

Given user preference vector $\mathbf{u}$, in each evaluation session, the hidden prominent aspect is sampled by $k \sim \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}}$. Evaluation on the prominent aspect is magnified by $\exp \theta$. Evaluation on a non-prominent aspect is denoted as $\mathbf{q}_k - \mathbf{b}_{u,k}$. Thus, when the aspect k is chosen, the evaluation of user u on q is $\mathbf{q}_k \exp \theta + \sum_{k' \neq k} (\mathbf{q}_{k'} - \mathbf{b}_{u,k'})$. The prediction is generated across all possible hidden prominent aspects. This gives us the following non-compensatory versions of rating prediction formulas.

Non-compensatory Matrix Factorization: MF-N

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

$$(9)$$

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

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \frac{\exp(\sum_{\mathbf{p} \in R(u)} \mathbf{p}_k)}{\sum_{k'} \exp(\sum_{p \in R(u)} \mathbf{p}_{k'})} \quad (10)$$
$$[\mathbf{q}_k \exp \theta + \sum_{k' \neq k} (\mathbf{q}_{k'} - \mathbf{b}_{u,k'})].$$

Non-compensatory Local Low Rank Matrix Approximation: LLORMA-N uses the same decomposition for each sub-matrix.

$$\hat{\mathbf{X}}_{u,q} = \sum_{t=1}^{S} \sum_{k} \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}} \frac{K((\mathbf{u}_t,\mathbf{i}_t),(\mathbf{u},\mathbf{q}))}{\sum_{s=1}^{S} K((\mathbf{u}_s,\mathbf{i}_s),(\mathbf{u},\mathbf{q}))} \quad (11)$$
$$[\mathbf{q}_{t,k} \exp \theta + \sum_{k' \neq k} (\mathbf{q}_{t,k'} - \mathbf{b}_{u,k'})]$$

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

The model parameters, including $\mathbf{U}, \mathbf{V}, \theta, \mathbf{b}$ are learnt during the training process. To infer the parameters, we employ the standard gradient descent framework.

Non-Compensatory Ranking Models

Non-compensatory Thurston Model. 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-specific aspect specific threshold $\mathbf{b}_{u,k}$ cancels between $\hat{\mathbf{X}}_{u,p}$ and $\hat{\mathbf{X}}_{u,q}$. Thus the model parameters consist only $\Theta = \{\mathbf{U}, \mathbf{V}, \theta\}$.

The inference of BPR-N is straightforward. As the loss function $\mathcal{L}(\Theta)$ consists of a term for the log-likelihood which is denoted by $\mathcal{L}'$, and a penalty term $\lambda_U \|U\|_2^2 + \lambda_V \|V\|_2^2$, the stochastic gradient descent (SGD) in inference can also be expressed by two terms $\frac{\partial \mathcal{L}'}{\partial \Theta}$ and $\lambda_U U + \lambda_V V$. Furthermore, $\frac{\partial \mathcal{L}'}{\partial \Theta} = \sum_u \sum_{p \succ_u q} \frac{\partial \mathcal{L}'}{\partial \Delta \hat{\mathbf{X}}_{u,p,q}} \frac{\partial \Delta \hat{\mathbf{X}}_{u,p,q}}{\partial \Theta}$, where $\Delta \hat{\mathbf{X}}_{u,p,q} = \hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}$. The same procedure is applied to infer parameters in the non-compensatory versions of FS-BPR (Zhao et al. 2018) and LCR (Lee et al. 2014).

Non-compensatory Bradley-Terry Model. Finally we propose BT-N, the non-compensatory version of BT (Hu and Li 2016). In order to treat prominent and non-prominent aspects differently, we define the probability of preference $p \succ_u q$ as the product of factor-wise comparisons. Intuitively p outranking q suggests that p is significantly better on the prominent aspect, and not too bad on non-prominent aspects. Inspired by a variant of BT model with ties (Hunter 2004), given a hidden prominent aspect $k \sim u_k$ sampled for the evaluation session, we define the factor-wise ranking probability on prominent aspect k as $\frac{\mathbf{p}_k}{\mathbf{p}_k + \theta \mathbf{q}_k}$, and ranking probability on non-prominent aspect k' as $\frac{\theta \mathbf{p}_{k'}}{\mathbf{q}_{k'} + \theta \mathbf{p}_{k'}}$. The overall prediction is aggregated over all possible hidden prominent aspect k:

$$Pr(p \succ_u q) = \sum_{k=1}^{K} \mathbf{u}_k \Big[\frac{\mathbf{p}_k}{\mathbf{p}_k + \theta \mathbf{q}_k} \prod_{k' \neq k} \frac{\theta \mathbf{p}_{k'}}{\mathbf{q}_{k'} + \theta \mathbf{p}_{k'}} \Big], \quad (12)$$

where $\mathbf{u}_k > 0, \sum_k \mathbf{u}_k = 1, \mathbf{p}, \mathbf{q} > \mathbf{0}$ and $\theta > 1$.

BT-N is also a combination of lexicographic rules and conjunction rules. (1) θ controls the strength of evaluation on the prominent aspect. Consider only the prominent aspect, the evaluation requires $p_k > \theta q_k, \theta > 1$ to generate a positive ranking $p \succ q$. Thus a larger value θ indicates that the user is more strict on the prominent aspect. When $\theta \to \infty$, the probability from non-prominent aspects approaches to 1, the users adopt lexicographic rules only. (2) The model sets the threshold in a dynamic manner. p is considered to be as good as q on a non-prominent aspect k', as long as $\theta p_{k'} > q_{k'}, \theta > 1$. An interpretation is that the user sets a minimal acceptance threshold for $p_{k'}$ based on the compared alternative $q_{k'}$, where the threshold is $q_{k'}/\theta$.

To infer the parameters $\mathbf{U}, \mathbf{V}, \theta$ 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

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

where t indicates the value obtained from the $t-$th round of SEM algorithm. In each M-step, we incorporate the MM bound proposed 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 to examine whether the non-compensatory versions outperform the original models on (1) rating data sets, (2) rankings generated from explicit rating feedback, and (3) rankings generated from graded implicit feedback. We also analyze the inferred parameters $\theta, \mathbf{b}$ in non-compensatory rules for further insights. The source codes are publicly available[1].

Comparative Results for Rating Prediction Models

Data Sets We use the standard benchmarks with user-item ratings. (1) Movielens[2]; (2) FilmTrust (Guo, Zhang, and Yorke-Smith 2013); and (3) CiaoDVD (Guo et al. 2014). 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 set 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, **Comparative Methods**. We compare the non-compensatory versions (with suffix "-N") with the original versions on three widely adopted rating prediction methods. (1) MF (Koren, Bell, and Volinsky 2009) in Equ. 2 v.s. MF-N in Equ. 9; (2) AMF (Koren 2008) in Equ. 3 v.s. AMF in Equ. 10, both without regularization terms; (3) LLORMA (Lee et

---

[1]https://github.com/XMUDM/Non-Compensatory

[2]https://grouplens.org/datasets/movielens/

Table 3: Statistics of Datasets with ratings

Dataset	#users	#items	#ratings	#pairs
Movielens	942	1,650	80,000	1,072,237
FilmTrust	1,235	2,062	35,497	623,516
CiaoDVD	2,665	14,280	72,665	2,478,836

al. 2013) in Equ. 4 v.s. LLORMA-N in Equ. 11. The non-compensatory and original versions adopt the same values for hyper-parameters. The model parameters to learnt are $\mathbf{U}, \mathbf{V}$ for compensatory models, and $\mathbf{U}, \mathbf{V}, \theta$ for non-compensatory models. In this section, we do not activate parameters $\mathbf{b}$ for computational efficiency, which is equivalent as setting $\mathbf{b}_{u,k} = 0$ for every u, k. We pose non-negative constraint on $\theta > 0$ in the inference. We stop the learning process either when the convergence is achieved (i.e. $\|\mathbf{U}^{t+1} - \mathbf{U}^t\| \leq 0.0001$) or when the algorithm reaches $MaxIter$ iterations. The same values of hyper-parameters are set for compensatory and non-compensatory models: for MF and MF-N $K = 10, \lambda_U = \lambda_V = 0.01, MaxIter = 1000$, for AMF and AMF-N $K = 5, MaxIter = 100$, for LLORMA and LLORMA-N $K = 5, S = 50, \lambda_U = \lambda_V = 0.001, MaxIter = 100$.

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; (2) NDCG: measures the accuracy of item ranking per user by the predicted ratings v.s. the actual ranking; (3) MRR: computes the reciprocal of the position of the item with the largest observed rating in the predicted ranking for each user. Results are averaged over all users.

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 in general 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 can be calculated 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.

Comparative Results for Ranking Models: Explicit Feedback

Data Sets. Next we evaluate the performance of models that target to ranking reconstruction. The datasets used are again Movielens, Filmtrust and CiaoDVD. For Movielens we use pre-processing step in (Hu and Li 2017), i.e. $\mathbf{X}_{u,p} \geq 4 \&\& \mathbf{X}_{u,q} \leq 2 \to p \succ_u q$. For Filmtrust and CiaoDVD, as these data sets are more sparse, we construct pair-wise ordering for each user between any higher rated item and

Table 4: Comparative rating prediction performances, 'Imp.' is the percentage of improvements of non-compensatory versions relative to the original models. Non-compensatory rules universally improve rating models.

Method	AUC	Imp.	NDCG	Imp.	MRR	Imp.
Movielens		(%)		(%)		(%)
MF	0.6729		0.6925		0.8300	
MF-N	0.7108	5.62	0.7166	3.48	0.8633	4.01
AMF	0.6901		0.7107		0.8747	
AMF-N	0.7017	1.68	0.7143	0.50	0.8769	0.26
LLORMA	0.7265		0.8734		0.7015	
LLORMA-N	0.7299	0.47	0.8999	3.03	0.7187	2.45
Filmtrust		(%)		(%)		(%)
MF	0.6507		0.5229		0.7011	
MF-N	0.6710	3.12	0.5241	0.23	0.7071	0.86
AMF	0.5971		0.5137		0.7411	
AMF-N	0.6159	3.15	0.5260	2.40	0.7649	3.22
LLORMA	0.6240		0.8596		0.7857	
LLORMA-N	0.6345	1.68	0.8684	1.02	0.8068	2.69
CiaoDVD		(%)		(%)		(%)
MF	0.7431		0.7949		0.8910	
MF-N	0.7903	6.34	0.8127	2.25	0.9154	2.74
AMF	0.6489		0.6612		0.8741	
AMF-N	0.7017	8.14	0.6892	4.24	0.8996	2.91
LLORMA	0.6752		0.7827		0.8267	
LLORMA-N	0.6845	1.38	0.7984	2.00	0.8384	1.42

Table 5: Non-compensatory rules generally improve ranking models on explicit feedback.

Method	AUC	Imp.	NDCG	Imp.	MRR	Imp.
Movielens		(%)		(%)		(%)
BT	0.6453		0.5329		0.8227	
BT-N	0.8511	31.89	0.5795	8.74	0.9256	12.51
BPR	0.7976		0.5674		0.8988	
BPR-N	0.8361	4.82	0.5761	1.53	0.9180	2.14
FSBPR	0.5048		0.5011		0.7524	
FSBPR-N	0.8272	63.86	0.5740	14.56	0.9136	21.42
LCR	0.7191		0.8555		0.9461	
LCR-N	0.736	2.35	0.8605	0.58	0.9515	0.57
FIlmtrust		(%)		(%)		(%)
BT	0.5405		0.5092		0.7702	
BT-N	0.6969	28.94	0.5446	6.95	0.8485	10.15
BPR	0.6412		0.5319		0.8206	
BPR-N	0.6729	4.94	0.5391	1.35	0.8364	1.93
FSBPR	0.4857		0.4968		0.7428	
FSBPR-N	0.6717	38.29	0.5388	8.47	0.8358	12.52
LCR	0.5977		0.9034		0.7511	
LCR-N	0.6144	2.79	0.9063	0.32	0.7635	1.65
CiaoDVD		(%)		(%)		(%)
BT	0.6063		0.5240		0.8031	
BT-N	0.9334	53.95	0.5981	14.1	0.9666	20.36
BPR	0.6344		0.5304		0.8172	
BPR-N	0.8987	41.66	0.5902	11.28	0.9493	16.17
FSBPR	0.7537		0.5574		0.8769	
FSBPR-N	0.8992	19.30	0.5903	5.91	0.9496	8.30
LCR	0.626		0.9408		0.7889	
LCR-NCR	0.6349	1.42	0.9451	0.46	0.7988	1.25

lower rated item, i.e. $\mathbf{X}_{u,p} > \mathbf{X}_{u,q} \rightarrow p \succ_u q$. The number of ranking pairs on each dataset is shown in Table 3.

Comparative Methods. We compare the non-compensatory versions with the original versions on four widely adopted ranking methods. (1) BT (Hu and Li 2016) which minimizes Equ 8 v.s. BT-N in Equ. 12, (2) BPR (Rendle et al. 2009) which optimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 2 v.s. BPR-N which optimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 9, (3) FSBPR (Zhao et al. 2018) which optimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 3 v.s. FSBPR-N which minimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 10, (4) LCR (Lee et al. 2014) which optimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 4 v.s. LCR-N which optimizes Equ. 6 with $\hat{\mathbf{X}}$ predicted by Equ. 11. The hyper-parameters are, $K = 5, MaxIter = 100$ for all models, for BPR and BPR-N $\lambda_U = \lambda_V = 0.3$, for FSBPR and FSBPR-N $\lambda_U = \lambda_V = 0.01$, for LCR and LCR-N $S = 50, \lambda_U = \lambda_V = 10^{-8}$. **Evaluation Metrics** include AUC, NDCG and MRR.

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

Comparative Results for Ranking Models: Implicit Feedback

In most recommender systems, users give not only 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 superior than an item q clicked in the same

---

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

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 non-compensatory 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. Unlike previous experiments, here 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. The hyper-parameters are, $K = 5, MaxIter = 100$ for all models, for BPR and BPR-N $\lambda_U = \lambda_V = 0.05$, for FSBPR and FSBPR-N $\lambda_U = \lambda_V = 0.01$, for LCR and LCR-N $S = 50, \lambda_U = \lambda_V = 10^{-6}$.

Evaluation Metrics. In addition to the aforementioned ranking evaluation metrics, i.e. AUC, NDCG and MRR, in order to evaluate the sessional ranking performance, we also adopt 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.

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. The improvement is particularly notable for BT model on less sparse data sets, such as Tmall-hybrid and Yoochoose. This is because a session of multiple ranking pairs is more informative, thus fixing the prominent aspect in a whole session is beneficial.

Analysis of Inferred Parameters

Finally, we analyze the values of inferred parameters $\theta, \mathbf{b}$ in non-compensatory models to gain some insights about the non-compensatory rules.

As the user-specific 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 split 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

---

^{4}http://2015.recsyschallenge.com

u over all aspects. Then we report the mean and standard deviation of σ_u over all users.

Strength of Non-compensatory Rules. We can see that the optimal value of θ is moderate, indicating the users adopt a combination of lexicographical rules and conjunctive rules.

Effect of user-specific Aspect-specific Threshold. We can see that the standard deviation is positive $\sigma(\mathbf{b}_u) > 0$, suggesting that the user-specific threshold for each aspect is significant different. Note that conventional matrix factorization links the same user-bias to all aspects. Our experimental result demonstrates that it is neccessary to employ aspect-specific threshold in non-compensatory rules.

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 universally boosts recommendation performance for a variety of rating prediction and ranking models.

References

Devooght, R.; Kourtellis, N.; and Mantrach, A. 2015. Dynamic matrix factorization with priors on unknown values. In *Proceedings of ACM SIGKDD Conference*, 189–198.

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 ASONAM Conference*, 540–547.

Guo, G.; Zhang, J.; and Yorke-Smith, N. 2013. A novel bayesian similarity measure for recommender systems. In *Proceedings of IJCAI Conference*, 2619–2625.

Hauser, J. R.; Ding, M.; and Gaskin, S. P. 2009. A critical review of non-compensatory and compensatory models of consideration-set decisions. In *Proceedings of the Sawtooth Software Conference*, 207–232.

Hu, J., and Li, P. 2016. Improved and scalable bradley-terry model for collaborative ranking. In *Proceedings of IEEE ICDM conference*, 949–954.

Table 7: Non-compensatory rules generally improve ranking performance on implicit feedback.

Dataset	Method	AUC	Improve(%)	NDCG	Improve(%)	MRR	Improve(%)	MAP	Improve(%)	Prec	Improve(%)
Tmall-single	BT	0.5304		0.2804		0.4870		0.4327		0.2778	
	BT-N	0.5400	1.82	0.2840	1.28	0.4948	1.61	0.4386	1.34	0.2801	0.84
	BPR	0.5181		0.2794		0.4854		0.4297		0.2767	
	BPR-N	0.5349	3.24	0.2848	1.92	0.4960	2.18	0.4401	2.41	0.2806	1.41
	FSBPR	0.5265		0.2824		0.4913		0.4350		0.2794	
	FSBPR-N	0.5389	2.35	0.2863	1.39	0.4988	1.53	0.4432	1.90	0.2818	0.87
	LCR	0.5231		0.8116		0.4093		0.3407		0.2430	
	LCR-N	0.5263	0.61	0.8141	0.31	0.4126	0.81	0.3436	0.85	0.2462	1.32
Tmall-hybrid	BT	0.5867		0.3015		0.5373		0.4929		0.2904	
	BT-N	0.6568	11.94	0.3279	8.75	0.5990	11.48	0.5527	12.13	0.3036	4.53
	BPR	0.6183		0.3183		0.5792		0.5318		0.2973	
	BPR-N	0.6460	4.48	0.3276	2.92	0.5990	3.41	0.5524	3.87	0.3030	1.94
	FSBPR	0.6334		0.3246		0.5916		0.5442		0.3026	
	FSBPR-N	0.6544	3.31	0.3309	1.94	0.6062	2.48	0.5603	2.95	0.3047	0.69
	LCR	0.5395		0.6778		0.4826		0.3975		0.2717	
	LCR-N	0.5636	4.47	0.6933	2.29	0.5140	6.51	0.4242	6.72	0.2858	5.19
Yoochoose	BT	0.6027		0.4734		0.7151		0.6361		0.4560	
	BT-N	0.7000	16.15	0.5160	8.99	0.7869	10.04	0.7084	11.37	0.4785	4.92
	BPR	0.6700		0.5065		0.7713		0.6895		0.4737	
	BPR-N	0.6920	3.28	0.5131	1.31	0.7812	1.29	0.7027	1.91	0.4771	0.74
	FSBPR	0.3272		0.3658		0.5062		0.4599		0.4006	
	FSBPR-NCR	0.6198	89.45	0.4822	31.83	0.7169	41.62	0.6448	40.22	0.4650	16.08
	LCR	0.7933		0.9748		0.8177		0.8104		0.7834	
	LCR-N	0.8012	1.00	0.9768	0.21	0.8334	1.92	0.8264	1.97	0.7975	1.80

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

Dataset	Movielens	FilmTrust	CiaoDVD
θ	0.608 ± 0.105	0.667 ± 0.016	0.773 ± 0.051
$\sigma(\mathbf{b}_u)$	0.0095 ± 0.0024	0.0095 ± 0.0023	0.0093 ± 0.0022

Hu, J., and Li, P. 2017. Decoupled collaborative ranking. In *Proceedings of the 26th International Conference on World Wide Web*, 1321–1329.

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 conference*, 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 *Proceedings of the 30th 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*, 85–96.

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

Pilászy, I.; Zibriczky, D.; and Tikk, D. 2010. Fast als-based matrix factorization for explicit and implicit feedback datasets. In *Proceedings of the 4th ACM conference on Recommender systems*, 71–78.

Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, 452–461.

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 4th ACM Conference on Recommender Systems*, 269–272.

Steck, H. 2015. Gaussian ranking by matrix factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, 115–122.

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

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*.