

# 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 evaluation are based on a weighted or summated score over different aspects. 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 by the user as the prominent aspect. (2) Evaluations over different aspects are non-compensatory, i.e. the user has higher standards with respect to the item's performance on the prominent aspect and is more tolerant to the item's performance on other aspects. We show how these general assumptions can be applied to a wide range of existing recommender systems, including the point-wise prediction models such as conventional Matrix Factorization (MF), Matrix Factorization with neighborhood (AMF++), and low-rank matrix approximation (LLORMA), and the pair-wise ranking models such as Bradley-Terry model (BTL) and Thurstonian model (e.g. BPR). We experimentally show that adopting non-compensatory rules can significantly improve performance of these point-wise prediction and ranking 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 product over multiple relevant aspects and compute a weighted or summated score for each product. Then they will select the product with the highest score. The result of such a procedure is that a good performance on one aspect of a product 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 that in most cases consumers make consumption related choices based on non-compensatory rules (Engel, Blackwell, and Miniard 1986). Non-compensate rules do not allow the shortcomings of a product to be balanced out by its attractive features. For example, the evaluation could be solely on the most important aspect and item's performance on other aspects does not affect the final judgement.

To the best of our knowledge, no previous work has been devoted to modeling non-compensatory rules in recommender systems. Our goal in this paper is to study this unexplored area. Our primary contribution is to model the assumptions in non-compensatory rules by (1) transforming user preferences and item features into the same  $K$ -dimensional hidden space as in most existing latent factor models, (2) picking a prominent aspect in each decision session, (3) and adopting different evaluation strategy on prominent and non-prominent aspects. We show that this strategy can be easily applied to a wide range of recommendation models, 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 ranking reconstruction models such as BTL model (Hunter 2004) and BPR style Thurstonian model (Rendle et al. 2009). We experimentally show that the non-compensatory version of these models significantly outperform 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

size previous research work on the basis of combinations of different rating approximation formulas and ranking aware 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 Recommendations Models

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  $\mathcal{L}(\mathbf{X}, \hat{\mathbf{X}})$ . We list the 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.

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

For simplicity we ignore the user specific or item specific bias (Koren, Bell, and Volinsky 2009). A massive amount of techniques have been proposed to extend Equ. 1 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.** It is possible to embed neighborhood information which is proved to be useful in traditional collaborative filtering strategies in latent factor models. Instead of directly modeling user preferences, 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.

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

where  $R(u)$  is the set of rated items for  $u$ . Note that AMF is 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:

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

$\mathbf{u}_t, \mathbf{q}_t$  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,  $K(\cdot)$  are arbitrary smooth kernels.

### Ranking Models

Another goal of recommendation research is to reveal the observed rankings. We here consider pair-wise rankings  $p \succ_u q$ . The pair-wise rankings can be generated from pre-processing the ratings, i.e.  $\mathbf{X}_{u,p} > \mu, \mathbf{X}_{u,q} < \mu$  (), or from explicit and implicit feedback, i.e.  $\mathbf{X}_{u,p} \neq 0, \mathbf{X}_{u,q} = 0$  (Rendle et al. 2009). To generate the probability of pair-wise rankings  $p(p \succ_u q)$ , we assume that 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 .

**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.

$$p(p \succ_u q) = \frac{\hat{\mathbf{X}}_{u,p}}{\hat{\mathbf{X}}_{u,p} + \hat{\mathbf{X}}_{u,q}} \quad (4)$$

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

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

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

- BTL model has been by far leveraged with MF prediction formula ();
- Thurstone model with standard matrix factorization prediction formula is first presented as BPR (Rendle et al. 2009), which maximizes the likelihood of Thurstonian modeling of MF 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 factorized neighborhood 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: LLORMA is utilized in the Thurstone model in (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 pairwise 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 propose 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. The decision rules can be naturally explained in the latent factor models. Non-compensatory rules include lexicographic, conjunction and disjunction rules. In a lexicographic rule the consumer first ranks the features. He/she then ranks the profiles using successively the first-ranked feature, breaking ties with the second-ranked feature. Under lexicographic rules, items are evaluated on the most important aspect. Under conjunctive and disjunctive rules, the consumer imposes requirements for minimally acceptable values on each aspect separately. The conjunctive and disjunctive rules are often used in conjunction with lexicographic rules.

Table 1: Illustrative example of non-compensatory rules

Item	Prominent aspect	Not prominent aspects	
<b>p</b>	0.8	0.3	0.3
<b>q</b>	0.7	0.2	0.2
<b>l</b>	0.6	0.3	0.3

We illustrate the non-compensatory rules using three toy item vectors **p**, **q**, **l** in Table. . Suppose the user’s minimal acceptable value is 0.3, then if the evaluation is solely based on lexicographic rules, the result will be  $p \succ_u q \succ_u l$ . If the user adopts conjunctive rules only,  $p, l$  will be picked. If the user adopts disjunctive rules only,  $p, q, l$  are all satisfying. If the user adopts both lexicographic rules and conjunctive rules,  $p \succ_u l$  and  $l$  will not be chosen.

Inspired by the psychological findings, we propose the following assumptions. (1) We assume that in each evaluation session, there is a prominent aspect. The choice of prominent aspect is dependent on the user preferences. (2) We assume that the evaluation on prominent aspect is different from the evaluations on 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

**MF-NCR** samples the hidden prominent aspect by  $\frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}}$ , and the prediction is generated across all possible hidden prominent aspects:

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

We use the parameter  $\theta$  to control the strength of prominent aspects, when  $\exp \theta \leftarrow \infty$  indicates that the user adopts lexicographical rules only.

**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.

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

### Non-Compensatory Ranking Models

In order to treat prominent and non-prominent aspects differently, we propose **BTL-NCR** which defines the probability of any ranking pairs  $p \succ_u q$  as:

$$p(p \succ_u q) = \prod_{k=1}^K \mathbf{u}_k \left[ \frac{\mathbf{p}_k}{\mathbf{p}_k + \theta \mathbf{q}_k} \prod_{k' \neq k} \frac{\theta \mathbf{p}_{k'}}{\mathbf{q}_{k'} + \theta \mathbf{p}_{k'}} \right]. \quad (9)$$

where  $\mathbf{u}_k > 0$ ,  $\sum_k \mathbf{u}_k = 1$  and  $\theta > 1$ . The interpretation of BTL-NCR is that, given a hidden prominent aspect  $k$ , the probability of user  $u$  ranking  $p$  over  $q$  is the multiplication of ranking probability over all aspects. Furthermore, the overall probability  $p(p \succ_u q)$  is largely affected by the performance divergence between  $p$  and  $q$  on the prominent aspect  $k$ . The parameter  $\theta$  controls the tolerance range of user  $u$  on other aspects. Because  $\theta > 1$ , as long  $p$  does not perform too badly on non-prominent aspects  $k' \neq k$ , i.e.  $\theta \mathbf{p}_{i,k'} > \mathbf{q}_{j,k'}$ , then the ranking probability  $p(p \succ_u q)$  will not be decreased.

Finally, the modification of Thurston model is straightforward. For example, if we use the Bayesian maximum posterior estimator as in BPR (Rendle et al. 2009), the loss function is defined as:

$$\mathcal{L} = - \sum_u \sum_{p \succ_u q} \ln \frac{1}{1 + \exp - [\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}]} - \lambda \|\Theta\|, \quad (10)$$

where  $\Theta$  is the set of all parameters and  $\hat{\mathcal{R}}$  is predicted by any NCR-version of rating prediction formulas.

## Experiments

## 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. 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.
- 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.
- 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*.