Reviewer 1

1,Rationale of non-compensatory conceptual model (figure 1)

Yes there is theoretical rationale for the model. Every component in Fig.1 is supported by study in consumer behavior. We combine and model the most commonly adopted non-compensatory rules: lexicographic and conjunctive rule. Lexicographic rule is defined as: items are first compared on the most important attribute. If item is superior on the attribute, it is selected. Therefore, in Fig.1 we select one prominent aspect based on user preference. Under conjunctive rules, cutoff points are set for each attribute. If item meets the cutoffs on all attributes, it is chosen. Therefore, we introduce the threshold parameter b on non-prominent attributes.

Yes we can have more than one prominent aspects. However, to obtain an elegant, general framework, abstraction and simplification is required. It may raise issues such as determining the number of prominent aspects. We will explore this direction in the future.

2,Some results are worse for the non-compensatory models

Thank you for pointing out the exception (only BPR-N is worse) in Table 5. We have to apologize here that it is a silly mistake. During rebuttal, we’ve double checked our logs. We also have re-run the BPR and BPR-N experiments. It turned out that when we were processing the results to calculate improvements using MS Excel, we mistakenly input one entry. The AUC of BPR on Filmtrust should be 0.6244 (instead of 0.7825 in our manuscript, which is BPR’s MRR result). Other entries are correct. Thus, BPR-N improves BPR by 7.75%. All non-compensatory models outperform original models.

We are very sorry for this unfortunate mistake. It is our duty to check everything before submitting. We will do our best to remove all errors including typos and grammar errors in revision. We will make our source codes and all results publicly available.

Reviewer 2

1,Are neighborhood models non-compensatory? they should be compared

Neighborhood models are **compensatory**. For example, in item-KNN, K neighboring items act as attributes of interest to the user, her own rating on each neighboring item is her evaluation on a relevant attribute. Item-KNN sums up evaluations on K relevant attributes, which is exactly the definition of compensatory rules. Following the same argument, user-KNN are also compensatory.

We did not compare with pure neighborhood models because latent factor models usually outperform pure neighborhood models. However, we did consider incorporating neighborhood information in latent factor models. We explained that AMF -- one of the earliest and most famous model of this kind– is compensatory. We showed that AMF can be adapted to adopt non-compensatory rules and achieve better results.

2,How to select prominent features

We agree that the selection of prominent aspect is crucial. That is why it is learnt from the training data. The prominent aspect (indexed by k) is treated as a hidden variable in each evaluation session. k is sampled according to the user preference, which is a model parameter that needs to be learnt. In BT-N model, k is explicitly inferred in a stochastic EM algorithm. In other models, we find it more convenient to integrate out k as in Equation 8~10.

3,Latent factor models already has at least 4 user-defined hyper-parameters. Having to tune 2 additional parameters theta and b may be a problem. What are the values of theta and b and how to obtain

We are sorry that we didn’t make it clear in the manuscript. Parameters theta and b are to be learnt, as well as user preferences U and item features V. Not defined or tuned as hyper-parameters. In revision, we can provide parameter updating equations if necessary.

Non-compensatory models do not have many additional parameters. (1)theta is one global parameter for all users and items. (2)b does not appear in ranking models, because it cancels between two items rated by the same user. (3)Non-compensatory rating models without b also greatly outperform compensatory models. As mentioned in the experiments, to reduce the number of parameters, we set b=0 for all users and aspects. This is equivalent as removing b from the model (see Equ. 8~10). Table 4 validates that, we can get satisfying improvements with the same number of parameters as compensatory models. (4)Performance of a full model with b is even better, which is predictable and not provided due to page limit. We can add it in revision.

The parameter analysis session analyzes the optimal values of theta and b learnt in a full model. The findings support our assumptions, i.e. the learnt theta is moderate, indicating that users adopt a combination of lexicographical and conjunctive rule. The not-small std of learnt b within each user indicates that introducing an aspect-dependent user bias term is helpful.

4,Is there a price to pay computationally in non-compensatory model

We design the models so that the same algorithmic framework can be used for both compensatory and non-compensatory models. Furthermore, one can accelerate rating models by removing parameters b and still obtain satisfying results. In that case, the difference of computational cost is subtle.

5,When are rankings observed?

Rankings are observed from (1)ratings, e.g. if u’s rating on item I is greater than u’s rating on item j then i>j for u. (2)graded implicit feedback,e.g. in a user session, if I is purchased, and j is clicked, then i>j in the session.

Reviewer 3

1,How to calculate AUC?

Yes. To compute AUC, the ground truth of binary classification is: a normalized rating above 0.5 is labeled positive and otherwise negative.

2,Equ. 6, 7 and 11

Thank you very much for pointing that out. In this manuscript, we follow the non-exponential form of Bradley-Terry model. Equation 7 should be Xui/(Xui + Xuj). It is worthy to mention that, to make sure Xui and Xuj are positive, in implementation we add non-negative constraints on user preferences and item features. Equation 11 is correct.