NEAR: a partner to explain any factorised recommender system

Sixun Ouyang  
Insight Centre for Data Analytics  
University College Dublin, Ireland  
sixun.ouyang@insight-centre.org

Aonghus Lawlor  
Insight Centre for Data Analytics  
University College Dublin, Ireland  
aonghus.lawlor@insight-centre.org

ABSTRACT
Many explainable recommender systems construct explanations of the recommendations these models produce, but it continues to be a difficult problem to explain to a user why an item was recommended by these high-dimensional latent factor models. In this work, we propose a technique that joint interpretations into recommendation training to make accurate predictions while at the same time learning to produce recommendations which have the most explanatory utility to the user. Our evaluation shows that we can jointly learn to make accurate and meaningful explanations with only a small sacrifice in recommendation accuracy. We also develop a new algorithm to measure explanation fidelity for the interpretation of top-n rankings. We prove that our approach can form the basis of a universal approach to explanation generation in recommender systems.

CCS CONCEPTS
• Information systems → Learning to rank; Recommender systems; Interpretation; • Computing methodologies → Feature selection;

KEYWORDS
Recommender systems, Learn to rank, Interpretation, Explanations

ACM Reference Format:  

1 INTRODUCTION
Latent factorisation methods continue to receive substantial attention in a long history in recommender system field [3, 8, 13, 16, 17]. At the same time, the process of explaining the resulting recommendations has also received considerable attention from both academia and industry. Rating prediction has been addressed through various approaches [1, 7, 10, 14, 15], however, the top-n ranking explanation is a tough problem which has not been thoroughly tackled.

On the other hand, general explaining recommender systems leverage the importance of observable attributes to make interpretations [12, 14, 15]. Nevertheless, they all aim to explain a particular domain, and are difficult to extend to other realms. As such, we target on providing explanations on latent factors, the basis of recommender systems, to form a universal foundation of recommendation explanations. Specifically, our method plays as the extensive model that joint interpretations into recommendation training to produce top-n ranked lists that provide optimal personalised explanations, where this procedure we named as explain training.

In this paper, we propose a robust method Native Explanation System for Agnostic Recommendation (NEAR), a method for training a recommender system with joint explanations. Our experiments show that by sacrificing a little precision, the joint system is able to supply accurate recommendations alongside convincing explanations. We also introduce a new explanation fidelity algorithm which has interesting applications for evaluating any model for explaining a ranked list.

2 RELATED WORK
From the early days of recommender systems, many works have explored the utility of explanations. Herlocker et al. [5] studied various interpretations leveraging distinct data types and visualisation styles on collaborative filtering tasks, and concluded that these approaches lead to users paying more attention to the value of recommendation explanations. Many works interpret recommendations in different ways. Vig et al. [15] used community tags to explain related recommendations. Costa et al. [7] indicated an alternative approach, machine-generated reviews, to offer explanations for recommendations. Hoeve et al. [14] outlined an algorithm that uses global point of interests of the factors to construct explanations, which is called LISTwise ExplaiNet (LISTEN). Although these methods successfully addressed many explanation approaches, there are only few works which attempt to fully explain a top-n recommendation. In this paper, we focus on joint explanations into recommendation training as a closed circuit and offer reliable explanations for top-n recommendations.
3 NATIVE EXPLAINER FOR AGNOSTIC RECOMMENDATION

NEAR consists of two parts: a training phase and an explain phase. The training phase, aforementioned explain training, is the primary section that we alternate between a training step of the recommender system and NEAR to ensure explainable and accurate prediction, while the explain phase is used to generate latent factor based interpretations for top-n rank. The full process is shown in Fig. 1. The training phase starts by disturbing the user embeddings which are the input to the factorisation model. Once the perturbation is applied, we use a factor selection method [14] to identify the most salient factor in the recommendation set, mimicking the process of a user selecting the item with the most relevant factor. In this way, the factors which are found to be the most salient by the factor selection method, are associated with more relevant items in the top-n recommendations. Similar to training phase, explain phase employ a leave one out disturb method [12] at first step. Then we feed the perturbed embeddings into lasso regression to get the weights of each factor. Because lasso leverages the importance of the factors by their weights, we denote these weights as the contributions of each factor.

4 EXPERIMENTS

Our experiments are conducted on 2 real-world review datasets from different domains: Movielens [4], and Beeradvocates [9], and on SVD++ [6], to evaluate against regular training.

4.1 Is NEAR reliable?

In this section, we measure the recommendation performance by Root Mean Square Error (RMSE) [2] on test set. Fig 2 illustrates the performance comparison of explain training and non-explain training. These results highlight a number of interesting features of NEAR. First, the non-explain training model performs slightly better than the with-explain method, which indicates a minor trade-off in accuracy by jointly learning the personalised explanations. On the other hand, the with-explain method converges much quicker and reach a similar minimum RMSE to the non-explain methods. This demonstrates that the with-explain methods are both accurate and precise and provide relevant and persuasive recommendations to users.

4.2 Can NEAR contribute to top-n explanation?

To explain a top-n list, we initially compute the overall contributions of each users’ factor through the explain phase. Then for each factor, we re-employ the leave one out method in explain phase to create a perturbed embedding, and further making a new top-n recommendation. To do so, these top-n is purely based on a specific factor, so that we can leverage how critical of a factor to the top-n list. We develop the explanation fidelity algorithm [11] which shown in Eq. 1. This fidelity algorithm counts the number of top-n instances where the average NDCG@k of top m factors is greater than that of other factors. The primary advantage is that we avoid bias in the results by incorporating weights of many factors instead of just one.

\[
F_{k, m} = \frac{\sum_{r=1}^{R} \text{NDCG}@k(f_r \in m) > \text{NDCG}@k(f_r \not\in m)}{R}
\]  

Evaluation on top-n explanation is demonstrated in Fig. 3, where the x-axis denotes factor ids which is sorted in decreasing order by their contributions, while the y-axis is the explanation fidelity. According to this, several observations can be made: (1) explanation fidelity decreased as m increased, which indicates NEAR returns the factors in the correct order; (2) explanation fidelity also declined along with increasing k. This is an expected outcome because it is impractical to maintain the same fidelity for different k values by the same size of factors; (3) NEAR is able to sustain explanation fidelity when trying to interpret longer top-k recommendation ranked lists, as there are only minuscule decrements between NDCG@5 and NDCG@10 on both of the experiments.

5 CONCLUSIONS

In this paper, we proposed NEAR, a robust explanation method for any latent factorised recommendation system. We introduced a cost function of explanation-training and argued this training method is crucial for personalisation. We showed that NEAR can present accurate and meaningful explanations without sacrificing a lot of precision in the recommendations. Concerning top-N explanation, we demonstrated a new interpretation fidelity algorithm and confirmed NEAR can produce the substantial foundation for explaining any top-N recommendation list. The NEAR method comprises many features which make it useful for general latent factor models, and future work will explore some extensions of the method, applying it to larger review domains, applying it to non-review based datasets, and further improving the evaluation.

6 ACKNOWLEDGMENTS

This work is supported by Science Foundation Ireland through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289.
REFERENCES