Towards Interpretation of Recommender Systems with Sorted Explanation Paths

Fan Yang†, Ninghao Liu†, Suhang Wang‡ and Xia Hu†
†Department of Computer Science and Engineering, Texas A&M University
‡College of Information Sciences and Technology, Penn State University
Emails: {nacoyang,nhliu43,xiahu}@tamu.edu, szw494@psu.edu

***Abstract*—Despite the wide application in recent years, most recommender systems are not capable of providing interpretations together with recommendation results, which impedes both deployers and customers from understanding or trusting the results. Recent advances in recommendation models, such as deep learning models, usually involve extracting latent representations of users and items. However, the representation space is not directly comprehensible since each dimension usually does not have any specific meaning. In addition, recommender systems incorporate various sources of information, such as user behaviors, item information, and other side content information. Properly organizing different types of information, as well as effectively selecting important information for interpretation, is challenging and has not been fully tackled by conventional interpretation methods. In this paper, we propose a post-hoc method called Sorted Explanation Paths (SEP) to interpret recommendation results. Specifically, we first build a unified heterogeneous information network to incorporate multiple types of objects and relations based on representations from the recommender system and information from the dataset. Then, we search for explanation paths between given recommendation pairs, and use the set of simple paths to construct semantic explanations. Next, three heuristic metrics, i.e., credibility, readability and diversity, are designed to measure the validity of each explanation path, and to sort all the paths comprehensively. The top-ranked explanation paths are selected as the final interpretation. After that, practical issues on computation and efficiency of the proposed SEP method are also handled by corresponding approaches. Finally, we conduct experiments on three real-world benchmark datasets, and demonstrate the applicability and effectiveness of the proposed SEP method.**

***Index Terms*—Recommender systems; Post-hoc interpretations; Heterogeneous information network**

I. Introduction

The past decade has witnessed the increasing deployment of recommender systems in various fields such as e-commerce websites [1], social networks [2] and review aggregation platforms [3]. In general, recommender systems aim to provide customers with the items that they are more likely to be interested in. Despite the great development of recommendation models in terms of the improving accuracy [4] and broader application scenarios [5], many recommender systems still face one challenging problem, i.e., their recommendation results are not interpretable. On one hand, the lack of interpretability makes it difficult for deployers to comprehensively understand the effectiveness and defects of their systems [6]. On the other hand, customers are not sufficiently motivated as they may not realize their essential demand or interests. This issue becomes prominent especially for those recommender systems that utilize latent factors as internal representations in decision making process [7]. The information of user preferences and item characteristics behind user-item ratings can be effectively encoded into the latent representations. However, the latent representations usually are largely indecipherable, thus making the recommendations generated from them become opaque.

Some preliminary work has been proposed to provide interpretation for recommendation models. In general, the interpretation schemes for recommendation can be divided into three categories: (1) discovering similar users as representatives for the target customer, according to their rating behaviors [8] or social connection [9]; (2) associating the target customer with relevant items to indicate the user interests [10]; (3) resorting to descriptive attribute information to understand the preferences of users and the characteristics of items, where the format of the interpretations could be both textual [11] and visual [12]. Existing methods simply design the interpretation scheme as a homogeneous set of objects (e.g., users, items or attributes) extracted from recommender systems or corresponding datasets. In general, among the three categories, it is hard to tell which one is superior compared with the other two. Thus, utilizing all three complementary interpretation schemes and generating interpretation adaptively would be beneficial in enhancing the quality and robustness of recommendation explanation.

Besides existing efforts, there are still several challenges to be solved towards designing a flexible and comprehensive interpretation method for recommender systems. First, many existing recommendation models map users and items into latent representation space, which is not directly comprehensible. Our goal is to design a model-agnostic interpretation method which does not specify how the latent space is constructed. Second, a structured interpretation scheme, which is capable of incorporating multiple types of objects, would be more desirable than the methods with homogeneous explanations. However, it is a nontrivial task to effectively organize different type of objects into a joint structure. Third, effective metrics and efficient sorting for explanations are challenging to be designed. The relevant evaluation and filtering need to be conducted according to some commonly accepted claims towards human understandability.

To tackle the aforementioned challenges, we develop a sorted explanation path (SEP) based method to achieve the post-hoc interpretability of recommender systems with latent representations. The proposed SEP method first builds a normalized weighted heterogeneous information network

978-1-5386-9159-5/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDM.2018.00082

(NW-HIN) to incorporate all available information together. Then, to obtain the interpretations for a specific user-item recommendation pair, SEP conducts the path searching by a revised depth-first algorithm, aiming to find all potential simple paths between the targeting user and the recommended item. With the mined explanation path set, the proposed SEP method sorts all the candidates by the unsupervised ranking method based on three designed heuristic metrics (i.e. *credibility*, *readability* and *diversity*). Further, the top-K interpretations are extracted according to relevant ranking scores. Besides, some practical issues of SEP are also handled for robustness and efficiency. The main contributions of this paper are summarized as follows:

- We design a post-hoc interpretation method called SEP, targeting general recommender systems with latent representations of users and items;
- We propose three heuristic metrics for interpretation evaluation (i.e. *credibility*, *readability*, *diversity*), and further use these metrics to filter potential candidates;
- We provide an efficient and effective approach to sort explanations comprehensively without human subjectivity;
- We test the proposed SEP method on three real-world datasets to evaluate its applicability and effectiveness.

II. Preliminaries

We first introduce the notations used in this paper. We use lowercase alphabets to denote variables (e.g. x), uppercase alphabets to denote scalars (e.g. X), boldface lowercase alphabets to denote vectors (e.g. $\mathbf{x}$), boldface uppercase alphabets to denote matrix (e.g. $\mathbf{X}$), and calligraphic uppercase alphabets to denote sets or entities (e.g. $\mathcal{X}$). Besides, for a matrix $\mathbf{X}$, its transpose is denoted as $\mathbf{X}^\intercal$ and its i-th row (or column) vector is represented as $\mathbf{x}_i$.

A. Latent Factor Model

Latent factor model is a very popular model-based recommendation approach in recent years, due to its high capability in capturing the correlation between users and items [13]. For example, matrix factorization is one of the common approaches belonging to this category [14]. Let $\mathbf{R}^{M\times N}$ be the rating matrix, where M is the number of users and N is the number of items. r_{ui} is the (u,i)-th element of $\mathbf{R}$, indicating the rating score of user u to item i. Matrix factorization decomposes $\mathbf{R}$ into two low-rank matrices $\mathbf{P}$ and $\mathbf{Q}$ with the following objective function:

$$\min_{\mathbf{P},\mathbf{Q}} \sum_{u=1}^{M}\sum_{i=1}^{N}(r_{ui}-\mathbf{p}_u\mathbf{q}_i^\intercal)^2 e_{ui} + \lambda_1\|\mathbf{P}\|_{\mathrm{F}}^2 + \lambda_2\|\mathbf{Q}\|_{\mathrm{F}}^2, \quad (1)$$

where $\mathbf{P}\in\mathbb{R}^{M\times D}$ is the user latent matrix with its u-th row $\mathbf{p}_u$ being the latent features of user u, and $\mathbf{Q}\in\mathbb{R}^{N\times D}$ is the item latent matrix with its i-th row $\mathbf{q}_i$ being the latent features of item i. $D \ll \min(M,N)$ is the latent dimension. e_{ui} represents the rating indicator that equals to 1 if user u rated item i and equals to 0 otherwise. λ_1, λ_2 are regularization coefficients. After training, user u's rating on item i is predicted as $\hat{r}_{ui}\approx\mathbf{p}_u\mathbf{q}_i^\intercal$. The proposed interpretation method uses latent matrices $\mathbf{P},\mathbf{Q}$ as part of the input.

B. Normalized Weighted Heterogeneous Information Network

Heterogeneous information network (HIN) is a network with multiple types of objects (nodes) and diversified relations (links). HIN enables representation of complicated real world systems, and empowers us to mine various knowledge from it. In our work, the links of HIN are associated with normalized weights (NW), so the HIN here is extended to NW-HIN, which is the information carrier for interpretation generation in our proposed method. The detailed definition of NW-HIN is given as below [15]:

Definition 1 (Normalized Weighted HIN): NW-HIN is defined as a graph $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{W})$ with a node type mapping $\tau:\mathcal{V}\to\mathcal{A}$ and a link type mapping $\phi:\mathcal{E}\to\mathcal{B}$, subject to $|\mathcal{A}|>1$ and $|\mathcal{B}|>1$. Each $v\in\mathcal{V}$ has a particular object type $\tau(v)\in\mathcal{A}$, and each $e\in\mathcal{E}$ has a particular relation type $\phi(e)\in\mathcal{B}$. Each relation has a normalized link weight $w\in\mathcal{W}$, which falls into the interval $[-1,1]$, indicating the connection strength between the corresponding two objects.

C. Problem Statement

We denote the given recommender system as $\mathcal{R}$ and the corresponding dataset as $\mathcal{D}$. The latent representations of users and items learned from $\mathcal{D}$ are denoted as $\mathbf{P}$ and $\mathbf{Q}$ respectively. The rating matrix is denoted as $\mathbf{R}$, and the content information is represented as the set $\mathcal{C}$. For a given recommendation pair between targeting user u and recommended item i, we aim to generate relevant interpretations for this recommendation result, with the aid of $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$ and $\mathcal{C}$. Specifically, in this paper, the interpretation is referred to a path set $\mathcal{K}$ in NW-HIN, where u and i are the end nodes for all relevant paths. In particular, a path $k\in\mathcal{K}$ in NW-HIN is defined as below.

Definition 2 (Path): A path k is a data structure defined on NW-HIN $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{W})$, and is typically in the format of $v_1 \overset{e_1}{\longleftrightarrow} v_2 \overset{e_2}{\longleftrightarrow} \cdots \overset{e_{l-1}}{\longleftrightarrow} v_l$ to indicate the composite relationship between node $v_1=u$ and $v_l=i$.

III. Methodology

In this section, we will formally introduce the proposed SEP method. The overall pipeline of SEP is illustrated in Fig. 1. Specifically, our proposed method involves four steps: (1) NW-HIN construction; (2) Explanation path mining; (3) Explanation path quantification; and (4) Unsupervised path sorting. The details of each step are introduced as below.

A. NW-HIN Construction

The purpose of constructing NW-HIN is to organize various types of information in a structured format. The flexibility of NW-HIN facilitates the subsequent procedures of explanation extraction. An example of the constructed NW-HIN is given in the upper part of Fig. 2. For generality, to construct NW-HIN, we consider three types of objects (i.e., HIN nodes), and four types of relations (i.e., HIN links). The set of object types is $\mathcal{A}=\{\mathcal{N}_u,\mathcal{N}_i,\mathcal{N}_a\}$, where $\mathcal{N}_u$, $\mathcal{N}_i$, $\mathcal{N}_a$ respectively denote the *user*, *item*, *aspect*. The set of relation types is $\mathcal{B}=\{\mathcal{L}_{uu},\mathcal{L}_{ii},\mathcal{L}_{ui},\mathcal{L}_{ia}\}$, where $\mathcal{L}_{uu}$ denotes *user-user similarity*, $\mathcal{L}_{ii}$ denotes *item-item similarity*, $\mathcal{L}_{ui}$ means *user-item strength*, and $\mathcal{L}_{ia}$ represents *item-aspect relevance*. The elements in $\mathcal{N}_a$

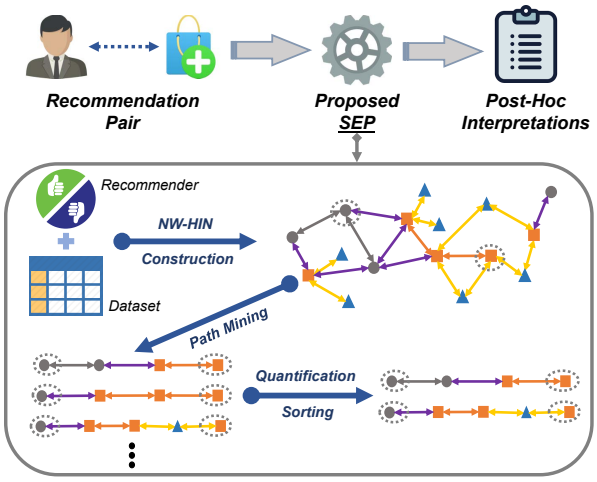


Fig. 1: Overall framework of the proposed SEP scheme.

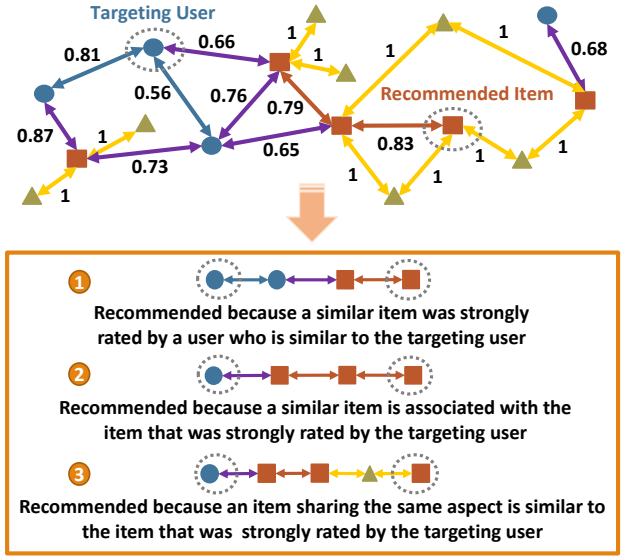


Fig. 2: Explanation path mining in a mock-up example.

depend on the specific applications. For example, in movie recommendations, $\mathcal{N}_a$ could include directors, genre, actors or some other properties that are related to items. Here, we do not consider $\mathcal{L}_{ua}$, i.e., the relation between $\mathcal{N}_u$ and $\mathcal{N}_a$, but it could be easily extended if relevant information source is available. The relation between two aspects is also not considered for simplicity in this work.

We assign weights to HIN links to quantify the relation strength between each pair of objects in the network. There are four types of relations defined in HIN. To measure the weight of links in type $\mathcal{L}_{uu}$, we employ the Pearson correlation coefficient [16] with $\mathbf{P}$ between the latent representations of two users. The Pearson correlation coefficient is also used to measure the weight of links in type $\mathcal{L}_{ii}$ with $\mathbf{Q}$. The weight of links in type $\mathcal{L}_{ui}$ is measured as:

$$w_{ui} = (r_{ui} - \bar{r}_{ui})/S \ ,$$

where S denotes the rating scale of $\mathcal{D}$, r_{ui} is the original rating in $\mathbf{R}$ scored by user u on item i, and $\bar{r}_{ui}$ indicates the average rating towards both user u and item i, which can be calculated by averaging all the ratings scored by u and scored on i. Furthermore, we measure the link weight for $\mathcal{L}_{ia}$ via a binary indicator shown as below:

$$w_{ia} = \begin{cases} 1 & \text{if } i \text{ contains } a \\ 0 & \text{if } i \text{ does not contain } a \end{cases} ,$$

which indicates the association between item i and aspect a.

At this point, the constructed NW-HIN has dense link connections (i.e., relations), which may require a large amount of storage when the data size is large. In practice, a threshold can be pre-set to filter the links whose weights are lower than the threshold, so that the obtained network could be relatively sparse. Generally, a larger construction threshold indicates a sparser NW-HIN with stronger relation links.

B. Explanation Path Mining & Quantification

After obtaining the NW-HIN, the interpretation of the given user-item pair is constructed as a group of Explanation Paths (EP) between the user object $u \in \mathcal{N}_u$ and the item object $i \in \mathcal{N}_i$ in the network. Each link in the path corresponds to one type of relation in $\mathcal{B}$. The lower part of Fig. 2 shows some examples of EPs for a given recommendation pair in a mock-up NW-HIN. Conventional interpretation methods, such as user-based [8] and item-based [10] methods, can be regarded as a special case of EP, where each path contains only one intermediate object and the length is fixed as 2. In these cases, the heterogeneous information sources in recommender systems are not fully utilized. For the proposed SEP method, to limit the number of valid paths, as well as to keep each path concise, we set the maximum length of each EP as 5 [17].

Given the NW-HIN, we employ a revised depth-first search algorithm [18] to mine the candidate EPs. To keep the path mining process efficient, we perform network pruning, so that only the partial network region around the targeting user and recommended item is preserved. Specifically, we only keep the nodes that are interconnected with both targeting user and recommended item within the maximum length 5. The search algorithm is further run on the pruned NW-HIN, to avoid EP mining on irrelevant network regions.

Nevertheless, it is not the case that all the mined EPs have good qualities for interpretation. To further evaluate the qualities of EPs, relevant criteria or metrics are needed for EP quantification. Specifically, we design three heuristic metrics for the EP quality evaluation, i.e., *Path Credibility*, *Path Readability* and *Path Diversity*. Each evaluation metric corresponds to one part of the properties for mined EPs. We introduce each metric, respectively, as follows.

Path Credibility: The credibility of an EP measures the overall link weight of the path. Typically, a higher weight for EP indicates a greater credibility of the corresponding interpretation. As we discussed in the NW-HIN construction part, different relations have different link weights, which actually signifies their disparate strength in delivering interpretations.

Therefore, referring to the social estimation method [19], the credibility of EP k is defined as:

$$Q^C(k) = \prod_{l=1}^{l_k} w_l, \tag{2}$$

where l_k denotes the length of EP k, and w_l represents the link weight of relation l on EP k.

Path Readability: The readability of an EP measures the legibility of corresponding explanations, which is also referred as the understandability of interpretation [20]. In reality, there are multiple factors that may affect the comprehension of interpretations [21]. Specifically, in our scenario, the readability of EP k is defined as follows:

$$Q^R(k) = 1 \Big/ \sqrt{l_k \cdot t_k^n}\,, \tag{3}$$

where t_k^n denotes the total number of object types for intermediate nodes in EP k. As shown in Eq. (3), the readability of an EP is inversely proportional to the path length and the number of intermediate object types, which matches our common senses in reality.

Path Diversity: The diversity of an EP reflects its variety in relation types among all the links. According to some studies in psychological science [22], a key factor affecting human's comprehensions and decisions is the information diversity. In general, human beings are more efficient to comprehend a relation chain with diverse information, and are more likely to accept the decision led by this chain. Thus, diversity is another perspective to be considered when evaluating the quality of EPs. Referring the work [23], the path diversity of EP k can be calculated as below:

$$Q^D(k) = \log_{l_k+1}(t_k^e + 1), \tag{4}$$

where t_k^e denotes the total number of relation types in EP k. From Eq. (4), we can see that the path diversity equals to 1 when $t_k^e = l_k$, and equals to 0 when $t_k^e = 0$.

After quantifying credibility, readability and diversity for EPs, we can denote an EP k as a 3-dimensional column vector $\mathbf{k}$, i.e., $\mathbf{k} = [Q^C(k), Q^R(k), Q^D(k)]^\intercal$. By utilizing the vectorized representations, it is convenient to further select out EPs with better qualities.

C. Unsupervised Path Sorting

In practice, although we set an upper bound to the length of EPs, there could still be a large number of available EPs for a given recommendation pair, due to the large size of the constructed NW-HIN. To effectively deliver interpretation results, we want to select a small set of EPs with good qualities instead of using all mined EPs. Generally, in the selection process, the top ones are expected to have higher credibility, readability and diversity. To achieve this, our proposed SEP method sorts all the EP candidates, and extract the top-K paths as our final interpretation.

Sorting EPs is a non-trivial task, since the three metrics we defined can possibly affect each other. For example, an EP with higher credibility could be less readable, and an EP with higher diversity also could be less credible. Thus, when sorting paths, a comprehensive score is needed to incorporate all of three metrics simultaneously. One straightforward way is to calculate the weighted sum over different metrics, and then select paths with top summation scores. However, the summation requires weights assignment to different metrics, usually determined by human, which makes it very subjective in different applications. To this end, in the proposed SEP method, we sort all the EPs in an unsupervised way, and learn the ranking function solely based on the vectorized EPs.

Instead of directly learning the ranking function, we utilize its inverse function, i.e. principal curve [24], to obtain the approximate mapping from the EP vector space to the ranking score. Due to some geometric properties [25], Bezier curves is commonly used for smooth function modeling in many applications. Referring to work [26], we employ the ranking principal formulated with cubic Bezier curves to achieve the comprehensive path sorting. Based on the definition of Bezier curves, our cubic ranking principal is formulated as follows:

$$\mathbf{f}(s) = \sum_{h=0}^{3} b_h^3(s)\mathbf{y}_h, \quad s \in [0,1], \tag{5}$$

where s denotes the ranking score, $b_h^3(s) = \frac{6(1-s)^{3-h}s^h}{h!(3-h)!}$ is the Bernstein polynomials, and $\mathbf{y}_h \in \mathbb{R}^3$ $(h = 0,1,2,3)$ indicates the control point of curve. In the matrix form, the cubic ranking principal curve is

$$\mathbf{f}(s) = \mathbf{Y}\mathbf{T}\mathbf{s}', \tag{6}$$

where

$$\mathbf{Y} = (\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3), \mathbf{T} = \begin{bmatrix} 1 & -3 & 3 & -1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{s}' = \begin{bmatrix} 1 \\ s \\ s^2 \\ s^3 \end{bmatrix}.$$

Given a candidate set of EPs mined from NW-HIN, i.e. $\mathcal{K} = \{\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_P\}$, we aim to learn the optimal control points $\mathbf{Y}$ of the principal curve and the corresponding ranking score s in Eq. (6). By minimizing the corresponding residual [24], our ranking problem can be further formulated as the following nonlinear optimization problem.

$$\min_{\mathbf{Y},\mathbf{s}} \quad L(\mathbf{Y},\mathbf{s}) = \sum_{p=1}^{P} \|\mathbf{k}_p - \mathbf{Y}\mathbf{T}\mathbf{s}^p\|^2 \tag{7}$$

$$\text{s.t.} \quad \mathbf{Y} \in [0,1]^{3\times 4}, \tag{8}$$

$$\mathbf{s} = (s_1, s_2, \cdots, s_P) \in [0,1]^{1\times P}, \tag{9}$$

$$\mathbf{s}^p = (1, s_p, s_p^2, s_p^3)^\intercal, \quad \forall p \in [1,P]. \tag{10}$$

Eq. (8) corresponds to the fact that our defined three metrics for EP k, i.e. $Q^C(k)$,$Q^R(k)$,$Q^D(k)$, all fall into the interval $[0,1]$. Eq. (9) is set to guarantee the existence of relevant derivatives. With the alternating optimization method, the local minimizer $(\mathbf{Y}^*, \mathbf{s}^*)$ could be obtained iteratively through the following two equations:

$$\arg\min_{\mathbf{Y}} \sum_{p=1}^{P} \|\mathbf{k}_p - \mathbf{Y}_{(t)}\mathbf{T}\mathbf{s}_{(t)}^p\|^2 = \mathbf{Y}_{(t+1)}, \tag{11}$$

$$2\left(\frac{\partial \mathbf{Y}_{(t+1)}\mathbf{T}\mathbf{s}^p}{\partial s^p}\right)^\intercal \left(\mathbf{k}_p - \mathbf{Y}_{(t+1)}\mathbf{T}\mathbf{s}^p\right)\Big|_{s^p = s_{(t+1)}^p} = 0, \tag{12}$$

where t denotes the iteration index. The existence of the feasible minimizer $(\mathbf{Y}^*, \mathbf{s}^*)$ has been demonstrated in [27].

To further represent the closed-form solution of local minimizer $(\mathbf{Y}^*, \mathbf{s}^*)$, we reformulate the problem in a matrix form. Organizing all the elements in $\mathcal{K}$ into the matrix $\mathbf{K}$, we have $\mathbf{K} = [\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_P]$. And the matrix $\mathbf{S}$ is formulated by combining all different $\mathbf{s}^p$ ($\forall p \in [1, P]$), i.e. $\mathbf{S} = [\mathbf{s}^1, \mathbf{s}^2, \cdots, \mathbf{s}^P]$. Then, the objective function in Eq. (7) can be rewritten as

$$L(\mathbf{Y}, \mathbf{S}) = \|\mathbf{K} - \mathbf{YTS}\|^2. \tag{13}$$

Letting the derivative of $L(\mathbf{Y}, \mathbf{S})$ equal to 0 regarding to $\mathbf{Y}$, we have

$$2\left(\mathbf{YTSS}^\intercal\mathbf{T}^\intercal - \mathbf{KS}^\intercal\mathbf{T}^\intercal\right) = 0. \tag{14}$$

Thus, the local optimizer $\mathbf{Y}^*$ can be explicitly indicated by

$$\mathbf{Y}^* = \mathbf{KS}^\intercal\mathbf{T}^\intercal\left(\mathbf{TSS}^\intercal\mathbf{T}^\intercal\right)^+ = \mathbf{K}\left(\mathbf{TS}\right)^+, \tag{15}$$

where $(\cdot)^+$ denotes the Moore-Penrose inverse computation. Given the intermediate result $\mathbf{S}_{(t)}$, we can update $\mathbf{Y}$ by $\mathbf{Y}_{(t+1)} = \mathbf{K}\left(\mathbf{TS}_{(t)}\right)^+$, and finally get $\mathbf{Y}^*$ after the convergence. With $\mathbf{Y}_{(t+1)}$ and $\mathbf{Y}^*$, $\mathbf{S}_{(t+1)}$ and $\mathbf{S}^*$ can be obtained from the solutions of s^p ($\forall p \in [1, P]$) to Eq. (12), which hardly has explicit general roots due to its high order.

With the aid of $\mathbf{S}^*$, considering the three metrics comprehensively, we can sort all EPs according to their ranking scores. After the sorting process, top EPs with good qualities will be selected out and further delivered to end-users for effective interpretations.

IV. Algorithm Implementation Design

Considering the robustness and efficiency of the SEP method in practice, we still have two obstacles ahead in algorithm implementation design. The first one lies in the computational issues caused by matrix inverse and ill-conditioned updates, and the second one results from the large size of the EP candidate set which largely slows the speed of sorting process. In this section, we employ two approaches to handle the practical issues, respectively.

A. Computation Relief

According to Eq. (15), we know that the control matrix $\mathbf{Y}$ is iteratively updated by $\mathbf{Y}_{(t+1)} = \mathbf{K}\left(\mathbf{TS}_{(t)}\right)^+$. For each iteration, there exists two computational issues: (i) the computation of $\left(\mathbf{TS}_{(t)}\right)^+$ is expensive, which could make the convergence much slower; (ii) the path representation matrix $\mathbf{K}$ could be ill-conditioned, which may lead to the instability of $\mathbf{Y}$ due to a small fluctuation $\Delta\mathbf{S}_{(t)}$. These two issues are directly related to the efficiency and robustness of the proposed SEP method. Here, in algorithm implementation, we use the Richardson iteration method [26] [28], to handle these two computational issues.

To relieve the computation from Moore-Penrose inverse calculation, we define a diagonal matrix $\mathbf{J}$, whose diagonal elements are the L_2 norm of columns in $(\mathbf{TS}_{(t)})(\mathbf{TS}_{(t)})^\intercal$. Then the update equation of $\mathbf{Y}$ can be expressed as:

$$\mathbf{Y}_{(t+1)} = \mathbf{Y}_{(t)} - \alpha^{(t)}\left[\mathbf{Y}_{(t)}(\mathbf{TS}_{(t)})(\mathbf{TS}_{(t)})^\intercal - \mathbf{K}(\mathbf{TS}_{(t)})^\intercal\right]\mathbf{J}^{-1},$$

where $\alpha^{(t)}$ is the parameter to guarantee the convergence of $\mathbf{Y}$. Typically, $\alpha^{(t)}$ is defined as $2\Big/(\lambda_{\max}^{(t)} + \lambda_{\min}^{(t)})$, where $\lambda_{\max}^{(t)}$ and $\lambda_{\min}^{(t)}$ respectively denote the maximum and minimum eigenvalue of $(\mathbf{TS}_{(t)})(\mathbf{TS}_{(t)})^\intercal$.

With the aid of the Richardson iteration method, we can avoid the expensive computation caused by Moore-Penrose inverse, and enhance the robustness of the iterative updates.

B. Sorting Acceleration

Another practical issue in algorithm implementation is the large size of the EP candidate set. With a large EP set $\mathcal{K} = \{\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_P\}$, the dimensions of $\mathbf{K}$ and $\mathbf{S}$ would also be very large, which may lead to the low speed of sorting process. For further acceleration, we use the Pareto frontier cull [29] to preprocess the EP candidate set to reduce its size.

The goal of the Pareto frontier cull is to filter out the EP candidates that are unlikely to be the top ones after sorting. Here, in our scenario, an EP candidate $\mathbf{k}_i$ is said to *dominate* another candidate $\mathbf{k}_j$ if the credibility, readability and diversity of $\mathbf{k}_i$ are all no smaller than those of $\mathbf{k}_j$. Mathematically, $\mathbf{k}_i$ dominating $\mathbf{k}_j$ can be indicated as:

$$\mathbf{k}_i \succeq \mathbf{k}_j \Longleftrightarrow \begin{cases} Q^C(k_i) \geq Q^C(k_j) \\ Q^R(k_i) \geq Q^R(k_j) \\ Q^D(k_i) \geq Q^D(k_j) \end{cases},$$

where the three inequalities do not take the equal sign simultaneously. If an EP cannot be dominated by any other candidates, then this EP candidate is called the *non-dominated* EP. The ultimate goal of the Pareto frontier cull is to find the *non-dominated* EPs, and further feed them into the path sorting process. In this way, the number of EP candidates fed into the sorting process would be significantly reduced, and it can save lots of efforts in ranking these *dominated* EPs.

C. Algorithm Steps and Time Complexity

By now, it is sufficient for us to summarize the overall steps of the proposed SEP method. Generally, the inputs of SEP are a pretrained recommender system $\mathcal{R}$ with latent factors $\mathbf{P}, \mathbf{Q}$, dataset $\mathcal{D}$, targeting user u, recommended item i and some threshold values δ, ξ for NW-HIN construction as well as convergence determination. The outputs of SEP are the top-K EPs for the given recommendation pair. The specific steps of the SEP method are illustrated by **Algorithm 1**.

In Algorithm 1, the time complexity of Line 1 and Line 2 together is $\mathcal{O}(|\mathcal{V}| + |\mathcal{E}|)$, where $\mathcal{V}$ and $\mathcal{E}$ respectively denote the node set and link set of the NW-HIN. The node/link construction takes $\mathcal{O}(1)$ time per operation and depth-first search needs $\mathcal{O}(|\mathcal{V}| + |\mathcal{E}|)$ time in total. From Line 3 to Line 8, $\mathcal{O}(3|\mathcal{K}| + |\mathcal{K}|) = \mathcal{O}(|\mathcal{K}|)$ time is needed for quantification and filtering, where $\mathcal{K}$ is the EP candidate set mined from the NW-HIN. For the sorting part, from Line 9 to Line 14, it needs time $\mathcal{O}(|\mathcal{K}| + 3 \times 4) = \mathcal{O}(|\mathcal{K}|)$, depending on the size of $\mathbf{Y}$ and $\mathbf{S}$. Thus, the overall time complexity of the proposed SEP method is $\mathcal{O}(|\mathcal{V}| + |\mathcal{E}| + |\mathcal{K}|)$.

Algorithm 1: Proposed SEP method

Input: $\mathcal{R}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathcal{D}$, u, i, δ, ξ.
Output: top-K EPs between u and i

```
Construct NW-HIN for R based on P, Q, D with δ;
Mine the EP candidate set K in NW-HIN with u and i;
for each EP k ∈ K do
    Calculate the credibility of EP k by Eq. (2);
    Calculate the readability of EP k by Eq. (3);
    Calculate the diversity of EP k by Eq. (4);
Formulate the EP representation matrix K;
Employ Pareto frontier cull on K;
Initialize the control matrix Y and score matrix S;
while ΔL(Y, S) > ξ do
    Update Y using the Richardson iteration method;
    Estimate the solutions of Eq. (12) for S;
    if ΔL(Y, S) < 0 then
        break;
Sort all the EPs according to ranking scores;
Export the top-K EPs as the final interpretations.
```

TABLE I: Dataset information in experiments.

Datasets	#Users	#Items	#Aspects	#Ratings	Scale
ML-100K	943	1,682	19	100,000	(1, 5)
ML-1M	6,040	3,900	18	1,000,209	(1, 5)
GB-10k	53,424	10,000	5,841	5,976,478	(1, 5)

V. Experiments

In this section, we evaluate the performance of the proposed SEP method on three real-world datasets. With the experiments conducted in this paper, we aim to answer three key questions as follows.

- How is the ability of the proposed SEP method in providing relevant interpretations, compared with the conventional and state-of-the-art methods?
- How much influences do the interpretations, generated from the SEP method, have to the given recommender system regarding to specific recommendation pairs?
- How do the defined metrics, including credibility, readability and diversity, affect the interpretations generated from the proposed SEP method?

A. Three Real-World Datasets

We use 3 publicly available benchmark datasets for recommender systems. Each dataset contains both rating information and relevant content information. For different datasets, we may use different attributes as our content information according to specific applications. The statistics of the 3 datasets are given in Table I.

- **MovieLens-100K**[1]: ML-100K is a popular benchmark dataset for movie recommendation. It contains a lot of content information besides ratings. In our experiments, we use *movie genre* attributes as the aspect objects, assuming that users may like the movies of certain genres.
- **MovieLens-1M**[2]: ML-1M is another benchmark dataset for movie recommendation. Similarly, we use *movie genre* as the aspect objects with the same assumption.
- **GoodBooks-10K**[3]: GB-10K is a book recommendation dataset. Besides ratings in the dataset, we employ *book author* attributes as the aspect objects, with the assumption that users may like the books by certain authors.

B. Baseline Methods

We compare the SEP method with several baseline methods introduced as below:

- **UBI** (User-Based Interpretation) [8]: UBI is a conventional interpretation method which matches the targeting user with similar users as the interpretation results.
- **IBI** (Item-Based Interpretation) [10]: IBI is another conventional interpretation method which generates interpretation as the closely associated items given the recommended one.
- **EMF** [30]: EMF is one of the state-of-the-arts for interpreting recommender systems. Specifically, this method was proposed targeting recommendation models based on matrix factorization. The generated interpretations from EMF are typically the neighbor-based explanations.
- **UniWalk** [31]: UniWalk is another state-of-the-art method for interpretable recommender systems. This method was also proposed based on matrix factorization models. The interpretations are generated by weighted random walks.
- **KNN-U** (K-Nearest Neighbors for Users) [32]: KNN-U selects the nearest users given the targeting one in the representation space of users.
- **KNN-I** (K-Nearest Neighbors for Items): KNN-I picks the nearest items given the recommended one in the representation space of items.

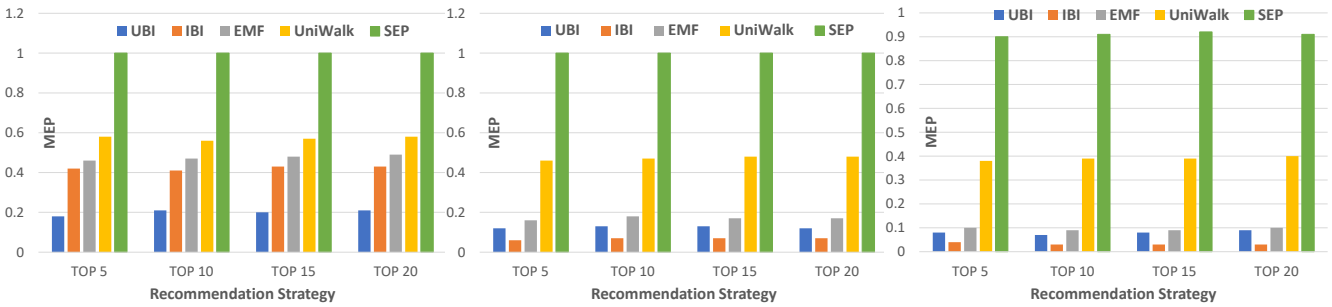
C. Experiment Settings

In our experiments, the given recommender system $\mathcal{R}$ is constructed based on the Non-negative Matrix Factorization (NMF) model, and the latent representation dimension D is set as 100. We apply the three datasets(i.e. ML-100K, ML-1M and GB-10K) for method evaluation. In order to have reproducible recommendation results for interpretation, we split each dataset where 80% data instances are used for training and 20% data instances are used for testing. Also, to guarantee that all methods run under the same conditions, we use the same threshold for network constructions. Specifically, relation links of $\mathcal{L}_{uu}, \mathcal{L}_{ii}, \mathcal{L}_{ui}$ are constructed if their link weights are above the 95-th percentile within the same relation type, and all $\mathcal{L}_{ia}$ links are built without thresholds. This setting is different from work [30] and [31], where the thresholds for network constructions are both set as 0. For the baseline method UniWalk, we use $\mathcal{L}_{uu}$ relation links instead of user social links during the network construction, since the applied datasets do not include social information. Also, we set the convergence threshold ξ in the SEP method as 10^{-3}. Finally,

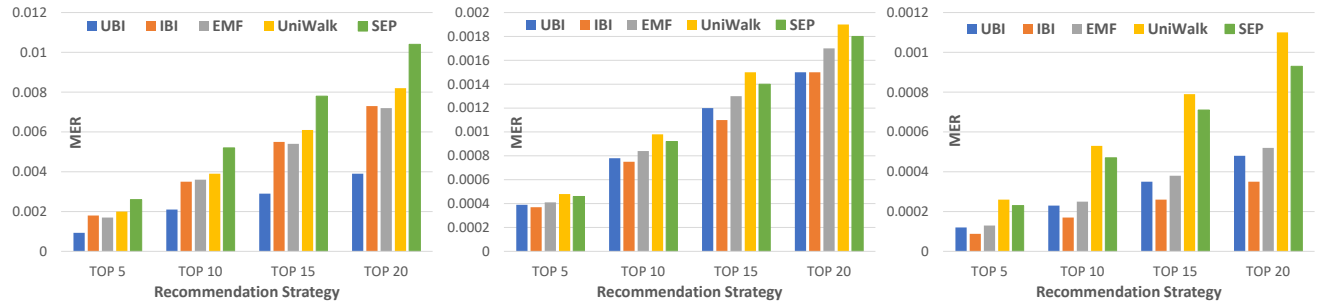
[1] https://grouplens.org/datasets/movielens/100k/

[2] https://grouplens.org/datasets/movielens/1m/

[3] https://github.com/zygmuntz/goodbooks-10k/



(a) MEP performance in ML-100K. (b) MEP performance in ML-1M. (c) MEP performance in GB-10K.



(d) MER performance in ML-100K. (e) MER performance in ML-1M. (f) MER performance in GB-10K.

Fig. 3: Applicability comparison among different methods in different datasets.

all of our experiments are conducted under Intel(R) Core(TM) i7-6850K CPU @3.60GHz with 128 GB memory.

D. Applicability of SEP Method

In this part, we evaluate the applicability of the SEP method compared with the baseline methods. The motivation behind is that, for some recommendation pairs, some interpretation methods are very likely to generate NULL as output (e.g., UBI returns nothing if it fails to find any similar user who has purchased the recommended item). We apply two metrics in our experiments, i.e. Mean Explainability Precision (MEP) and Mean Explainability Recall (MER), which are used in [30] and [33]. Specifically, MEP and MER are defined as follows.

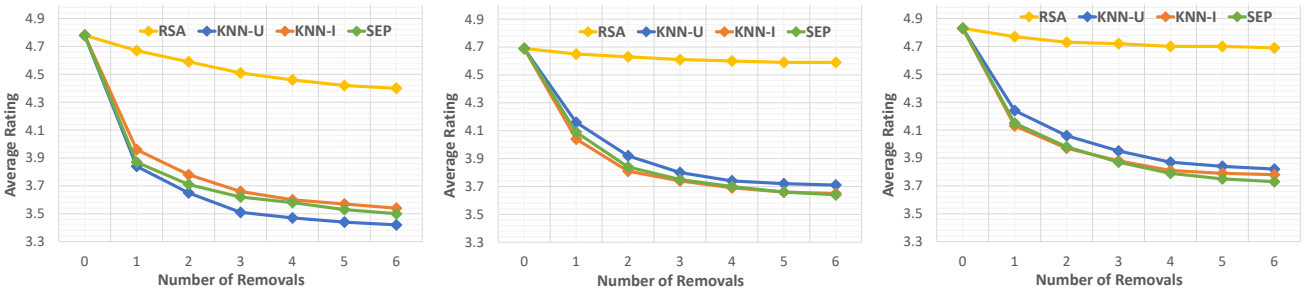
$$\text{MEP} = \sum_{u \in \mathcal{U}} \frac{|\mathcal{I}_u^{ir}|}{|\mathcal{I}_u^r|} \Big/ |\mathcal{U}| \,, \quad \text{MER} = \sum_{u \in \mathcal{U}} \frac{|\mathcal{I}_u^{ir}|}{|\mathcal{I}_u^i|} \Big/ |\mathcal{U}| \,,$$

where $\mathcal{U}$ indicates the sampling user set, $\mathcal{I}_u^{ir}$ is the interpretable recommended item set for user u, $\mathcal{I}_u^r$ denotes the recommended item set for user u, and $\mathcal{I}_u^i$ is the interpretable item set for user u. In our experiments, $|\mathcal{U}|$ is set as 10, $\mathcal{I}_u^r$ is determined by the recommendation strategy, and $\mathcal{I}_u^i$ depends on the specific interpretation method. Specifically, $|\mathcal{I}_u^r|$ equals to K with the top-K recommendation strategy, and $|\mathcal{I}_u^i|$ equals to the total number of items which have at least one explanation to user u. In general, the larger MEP and MER values correspond to the better applicability of methods.

The experiment results are shown in Fig. 3. Here, we compare the MEP and MER performance of the proposed SEP with the baseline methods on three different datasets. From Fig. 3(a) to 3(c), we observe that the proposed SEP method significantly outperforms the other four methods on MEP metric, which essentially means that almost every recommendation pair from $\mathcal{R}$ can be interpreted by the SEP method. From Fig. 3(d) to 3(f), as for the MER metric, we observe that the proposed SEP method slightly outperforms the other four methods in the small dataset ML-100K, and gets the competing performance as well in the larger datasets, where UniWalk ranks first in ML-1M and GB-10K. Overall, it is observed that SEP has a significant enhancement in MEP performance and has a competitive MER performance in most cases. Comprehensively, the proposed SEP method is demonstrated to have strong applicability in providing post-hoc interpretations for recommender systems.

E. Effectiveness of SEP Method

In this part, we demonstrate the effectiveness of the proposed SEP method. Particularly, we aim to test whether our generated interpretations accurately reflect how recommender systems make decisions in pairing users with items. The main idea of evaluation is to utilize adversarial samples [34] [35]. Specifically, we knock out from training data the objects that appear in the interpretation results, and then retrain the whole system with the modified training data. With the new recommender system, for a specific recommendation pair, we further compare the new prediction score with the original one to see how it changes. We choose the recommendation pairs with high rating scores to be observed. If the rating score decreases significantly after training data modification, it indicates that the removed objects play important roles in score prediction of the given recommendation pair. In our experiments, we compare the SEP method with three baseline



(a) Average ratings in ML-100K. (b) Average ratings in ML-1M. (c) Average ratings in GB-10K.

Fig. 4: Average ratings after removing relevant objects in different datasets.

TABLE II: Effects of sorting metrics in different scenarios.

Datasets	**ML-100K**				**ML-1M**				**GB-10K**			
Scenario	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
Average Credibility	0.267	0.278	0.286	0.382	0.203	0.225	0.229	0.314	0.126	0.158	0.163	0.256
Average Readability	0.511	0.389	0.485	0.323	0.395	0.286	0.363	0.248	0.375	0.274	0.358	0.252
Average Diversity	0.830	0.797	0.743	0.656	0.842	0.816	0.721	0.623	0.832	0.819	0.725	0.619

methods including KNN-U, KNN-I and Random Selection Algorithm (RSA). The RSA involved here is a naive baseline method for user/item selection, where users and items are randomly selected to be knocked out. For the objects selected from the SEP method (only for $\mathcal{N}_u$ and $\mathcal{N}_i$), they are knocked out sequentially as they appeared in the top EPs, along with the number of removed object gradually increases. The objects from KNN-U and KNN-I are knocked out sequentially according to their distance from the targeting user and recommended item, respectively.

The experiment results are shown in Fig. 4. Here, we compare the average rating of 10 original top recommendation pairs before and after the corresponding modifications in different datasets. As shown in Fig. 4, the removal of interpretation objects from SEP significantly affects the prediction scores of the given recommendation pairs. As for the objects selected by KNN-U and KNN-I, due to their high correlations with the recommendation pairs in the representation space, the average rating score also largely decreases when we knock them out in training. We observe that the interpretation objects selected by SEP have the competing influences under the given recommender system, compared with the objects picked by KNN-U and KNN-I, which straightforwardly demonstrates the fact that the interpretations generated from the SEP method are reasonable and meaningful.

F. Effects of Sorting Metrics

In this part, we explore the effects of different metrics, including credibility, readability and diversity, on the generated EPs. For contrast, we consider four different scenarios listed as below:

- **Scenario I**: Interpretations are generated considering all metrics, i.e., credibility, readability and diversity;
- **Scenario II**: Interpretations are generated based on their credibility and diversity;
- **Scenario III**: Interpretations are generated according to their credibility and readability;
- **Scenario IV**: Interpretations are generated simply based on their credibility.

For each scenario, we pick out 10 recommendation pairs and generate the top-3 interpretations for each pair. We will compare the changes of credibility, readability and diversity for the generated EPs across different scenarios. The value of each metric is averaged over all the EPs we obtain.

Table II presents the experiment results of this part, indicating the changes of the average credibility, average readability and average diversity in different scenarios. Using the results of scenario I as the baseline, we can observe that the EPs generated in scenario II are less readable in average after we remove the readability metric in interpretation extraction. Similarly, in scenario III, the average diversity of the generated EPs significantly decreases, compared with scenario I, after we ignore the corresponding metric in extraction. In scenario IV, when ignoring both readability and diversity, the generated EPs are even less readable and diversified in average, although the average credibility is relatively higher. According to the results, we can observe that the proposed SEP method extracts interpretations in a more comprehensive way, and the credibility is somewhat sacrificed for acquiring more readable and diversified interpretations.

G. Case Study

To intuitively understand the interpretations generated by the proposed SEP method, we further give a case study here drawn from the GB-10K dataset. Fig. 5 shows a sampled recommendation pair in GB-10K and its corresponding top-3 EPs extracted by the proposed interpretation method. From

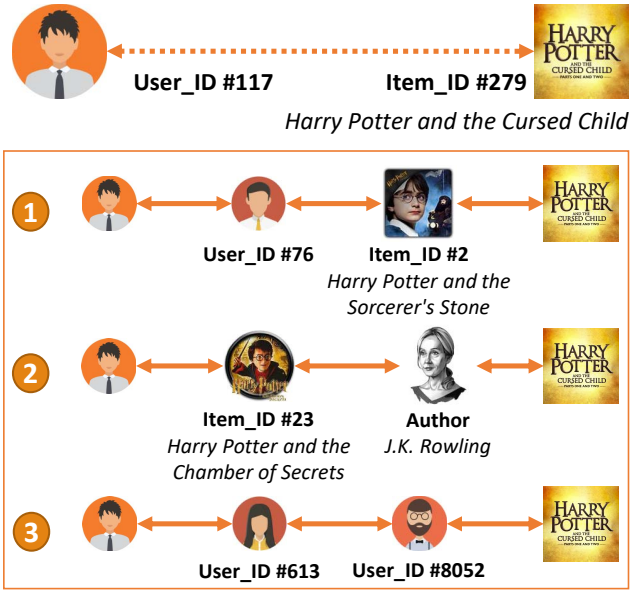


Fig. 5: A case study of recommendation interpretation.

the figure, we can observe that the generated interpretations from SEP are capable to provide explanations for user-item pairs in a flexible way, where different EPs associated with diversified semantics effectively help system administrators understand why particular items are recommended to them by the system. In this example, the book "Harry Potter and the Cursed Child" is recommended to user #117. The first reason provided by SEP is that a similar user #176 bought a book ("Harry Potter and the Sorcerer's Stone") which is highly associated with the recommended one. The second generated reason is that the book ("Harry Potter and the Chamber of Secrets") you purchased previously shares the same author (J.K. Rowling) with the recommended one. The third reason provided is that a similar user #8052 highly rated the recommended item. Besides, the generated interpretations are intuitively readable and diversified, where the relation chains are readily comprehensible and the relation types are multifarious. Comprehensively, the proposed SEP method can effectively and flexibly extract relevant interpretations for the specific recommendation pair.

VI. Related Work

A. Interpretability from Graph Construction

Since graph is a natural representation for objects and relations, lots of graph-based models have been employed in recommender systems to extract interpretations. Our work partially falls into this category. In [11], the authors proposed a tripartite graph model named TriRank to make top-K recommendations with relevant explanations. In TriRank, the interpretations are fully constructed based on the aspects extracted from user reviews. Similarly, the authors in [31] built a graph model, named UniWalk, to interpret recommendations through similar users and items by making weighted random walks. In [36], a user-item bipartite graph was built for explanations, and relevant interpretations were extracted by conducting co-clustering on the constructed graph. As a special case of graph, tree-based model has also been investigated for interpretable recommendations in [37], where relevant interpretations are generated through a designed attention network.

B. Interpretability from Topic Modeling

Another methodology for interpretability is topic modeling. Generally, the topic modeling is achieved with additional textual information, and the interpretations would be formulated through a series of topical words. In [38], through combining latent factor models and Latent Dirichlet Allocation (LDA), the authors designed the HFT model to help end-users understand the results by extracting important topics. With similar ideas, [39] modeled both user preference and item characteristics in a unified space, and further delivered relevant interpretations by the latent topics learned from user reviews. The authors in [40] proposed another topic modeling method, named sCVR, and the interpretations from sCVR are constructed using these top topics rated by users embedded in relevant viewpoints. Besides, there are also some recent work on topic modeling with probabilistic graphic models for interpretable recommendations, such as [41] and [42].

C. Interpretability from Knowledge Embedding

Knowledge embedding is a new methodology arisen recently for interpretable recommendations. In [43], the authors employed the rules and programming on knowledge graph to generate relevant interpretations. Particularly, with a personalized PageRank method, items and other graph objects are jointly ranked to produce corresponding recommendations and explanations. The authors in [44] used the knowledge graph embedding technique to extract explanations with a novel soft-matching algorithm, where the complicated relationships in knowledge base are briefly represented by embedding vectors. Creatively, in [45], the authors proposed an end-to-end system, named Ripple Network, to fully utilize the knowledge graph for interpretable recommendations. The interpretations generated from Ripple Network are typically a set of paths incorporated in the corresponding knowledge graph.

VII. Conclusion and Future Work

In this paper, we propose a post-hoc interpretation method called SEP for explaining the results of recommender systems. Specifically, by utilizing the latent representations from the recommendation model, together with the rating and attribute information, we first construct an unified information network. For each recommendation pair, a set of EPs are then extracted to indicate how the targeting user are related to the recommended item. Furthermore, through the unsupervised sorting process, the top-K EPs can be extracted from the path candidate set according to three designed metrics, i.e., credibility, readability and diversity. Experiments on three real-world datasets demonstrate the applicability and effectiveness of the proposed SEP method. The future extensions of this work may include exploiting textual reviews, incorporating structured knowledge base, and combining dynamics of recommendation scenarios into interpretation process.

ACKNOWLEDGMENT

The author(s) would like to thank the anonymous reviewers for their helpful comments and funding agencies for their generous supports. This work is, in part, supported by DARPA (#N66001-17-2-4031) and NSF (#IIS-1750074, #CNS-1816497). The views, opinions, conclusions and/or findings shown in this paper are those of the author(s), which should not be interpreted as representing the official views or policies of the Department of Defense, the U.S. Government or any funding agency.

REFERENCES

[1] A. Vieira and B. Ribeiro, "Recommendation algorithms and e-commerce," in *Introduction to Deep Learning Business Applications for Developers*. Springer, 2018, pp. 171–184.

[2] M. G. Campana and F. Delmastro, "Recommender systems for online and mobile social networks: A survey," *Online Social Networks and Media*, vol. 3, pp. 75–97, 2017.

[3] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang, "Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars)," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 193–202.

[4] B. Alhijawi, N. Obeid, A. Awajan, and S. Tedmori, "Improving collaborative filtering recommender systems using semantic information," in *Information and Communication Systems (ICICS), 2018 9th International Conference on*. IEEE, 2018, pp. 127–132.

[5] X. Zhou, M. Wang, and D. Li, "From stay to play–a travel planning tool based on crowdsourcing user-generated contents," *Applied Geography*, vol. 78, pp. 1–11, 2017.

[6] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," *arXiv preprint arXiv:1808.00033*, 2018.

[7] A. Datta, S. Kovaleva, P. Mardziel, and S. Sen, "Latent factor interpretations for collaborative filtering," *arXiv preprint arXiv:1711.10816*, 2017.

[8] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 2000, pp. 241–250.

[9] B. Wang, M. Ester, J. Bu, and D. Cai, "Who also likes it? generating the most persuasive social explanations in recommender systems." in *AAAI*, 2014, pp. 173–179.

[10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*. ACM, 2001, pp. 285–295.

[11] X. He, T. Chen, M.-Y. Kan, and X. Chen, "Trirank: Review-aware explainable recommendation by modeling aspects," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 1661–1670.

[12] X. Chen, Y. Zhang, H. Xu, Y. Cao, Z. Qin, and H. Zha, "Visually explainable recommendation," *arXiv preprint arXiv:1801.10288*, 2018.

[13] D. Agarwal and B.-C. Chen, "Regression-based latent factor models," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 19–28.

[14] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, 2009.

[15] Y. Sun and J. Han, "Mining heterogeneous information networks: a structural analysis approach," *Acm Sigkdd Explorations Newsletter*, vol. 14, no. 2, pp. 20–28, 2013.

[16] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[17] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna, "Four degrees of separation," in *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 2012, pp. 33–42.

[18] R. Sedgewick, *Algorithms in C, Part 5: Graph Algorithms, Third Edition*, 3rd ed. Addison-Wesley Professional, 2001.

[19] X. Lin, T. Shang, and J. Liu, "An estimation method for relationship strength in weighted social network graphs," *Journal of Computer and Communications*, vol. 2, no. 04, p. 82, 2014.

[20] H. Lakkaraju, S. H. Bach, and J. Leskovec, "Interpretable decision sets: A joint framework for description and prediction," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1675–1684.

[21] C. Tekfi, "Readability formulas: An overview," *Journal of documentation*, vol. 43, no. 3, pp. 261–273, 1987.

[22] J. S. Adelman, G. D. Brown, and J. F. Quesada, "Contextual diversity, not word frequency, determines word-naming and lexical decision times," *Psychological science*, vol. 17, no. 9, pp. 814–823, 2006.

[23] E. H. Simpson, "Measurement of diversity," *nature*, vol. 163, no. 4148, p. 688, 1949.

[24] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[25] G. Farin, *Curves and surfaces for computer-aided geometric design: a practical guide*. Elsevier, 2014.

[26] C.-G. Li, X. Mei, and B.-G. Hu, "Unsupervised ranking of multi-attribute objects based on principal curves," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 12, pp. 3404–3416, 2015.

[27] B.-G. Hu, G. K. Mann, and R. G. Gosine, "Control curve design for nonlinear (or fuzzy) proportional actions using spline-based functions," *Automatica*, vol. 34, no. 9, pp. 1125–1133, 1998.

[28] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.

[29] K. Deb, "Multi-objective optimization," in *Search methodologies*. Springer, 2014, pp. 403–449.

[30] B. Abdollahi and O. Nasraoui, "Explainable matrix factorization for collaborative filtering," in *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 5–6.

[31] H. Park, H. Jeon, J. Kim, B. Ahn, and U. Kang, "Uniwalk: Explainable and accurate recommendation for rating and network data," *arXiv preprint arXiv:1710.07134*, 2017.

[32] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[33] B. Abdollahi and O. Nasraoui, "Using explainability for constrained matrix factorization," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 2017, pp. 79–83.

[34] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[35] N. Liu, H. Yang, and X. Hu, "Adversarial detection with model interpretation," in *Conference on Knowledge Discovery & Data Mining*, 2018.

[36] R. Heckel, M. Vlachos, T. Parnell, and C. Dünner, "Scalable and interpretable product recommendations via overlapping co-clustering," in *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*. IEEE, 2017, pp. 1033–1044.

[37] X. Wang, X. He, F. Feng, L. Nie, and T.-S. Chua, "Tem: Tree-enhanced embedding model for explainable recommendation," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 1543–1552.

[38] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM Conference on Recommender Systems*. ACM, 2013, pp. 165–172.

[39] Y. Tan, M. Zhang, Y. Liu, and S. Ma, "Rating-boosted latent topics: Understanding users and items with ratings and reviews." in *IJCAI*, 2016, pp. 2640–2646.

[40] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, "Social collaborative viewpoint regression with explainable recommendations," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017, pp. 485–494.

[41] Y. Wu and M. Ester, "Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 199–208.

[42] K. Zhao, G. Cong, Q. Yuan, and K. Q. Zhu, "Sar: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015, pp. 675–686.

[43] R. Catherine, K. Mazaitis, M. Eskenazi, and W. Cohen, "Explainable entity-based recommendations with knowledge graphs," *arXiv preprint arXiv:1707.05254*, 2017.

[44] Q. Ai, V. Azizi, X. Chen, and Y. Zhang, "Learning heterogeneous knowledge base embeddings for explainable recommendation," *arXiv preprint arXiv:1805.03352*, 2018.

[45] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo, "Ripple network: Propagating user preferences on the knowledge graph for recommender systems," *arXiv preprint arXiv:1803.03467*, 2018.