

Explainable Reasoning over Knowledge Graphs for Recommendation

Xiang Wang^{1*}, Dingxian Wang^{2†}, Canran Xu², Xiangnan He^{1,3}, Yixin Cao¹, Tat-Seng Chua¹

¹School of Computing, National University of Singapore, ²eBay

³School of Information Science and Technology, University of Science and Technology of China

xiangwang1223@gmail.com, {diwang, canxu}@ebay.com, {xiangnanhe, caoyixin2011}@gmail.com, dcscts@nus.edu.sg

AAAI 2019

研究背景

先前的努力已经表明了将辅助数据结合到推荐系统中的重要性，例如结合用户信息和项目属性。最近，知识图（KG）引起了越来越多的关注，由于其全面的辅助数据：**项目的背景知识及其之间的关系**。它可以与用户项目交互无缝集成，更重要的是，通过探索知识图谱内的互连链接，用户和项目之间的连接反映了它们的基础关系，这些关系是对用户项目交互数据的补充。

来自KG的**额外用户项连接信息赋予推荐系统推理和解释能力**。推理揭示了交互背后可能存在的用户意图，提供了推荐背后的解释。因此，**如何在知识图谱中建模这种连接对于将知识注入推荐系统至关重要**。

先前对知识感知推荐的努力大致分为路径和嵌入方式。基于路径的方法引入了元路径以细化用户和项目之间的相似性。

然而，我们认为**元路径在推理KG时是低效的**：

- 很难指定路径的整体语义，特别是类似实体之间关系不同时；
- 无法自动发现和推理看不见的连接模式。

我们认为KGE正则化缺乏推理能力。

- 仅考虑实体之间的直接关系，而没有考虑**多跳关系**路径。
- 仅指导表示学习，但不推断用户的偏好。

主要工作

在这项工作中，提出通过推断用户对项目的偏好的原因来填补研究空白。在推理方面，我们期望我们的方法模拟实体的顺序依赖性和连接用户-项对的路径的复杂关系。在可解释性方面，我们希望我们的方法在推断用户兴趣时区分不同路径的不同贡献。

- 提出了一种名为**知识感知路径递归网络**（KPRN）的新解决方案，不仅计算实体和关系来生成路径表示，还基于**推断用户偏好的路径**执行推理。还**区分用于预测的不同路径的贡献**。
- 我们对两个数据集进行了大量实验，以验证我们的方法。

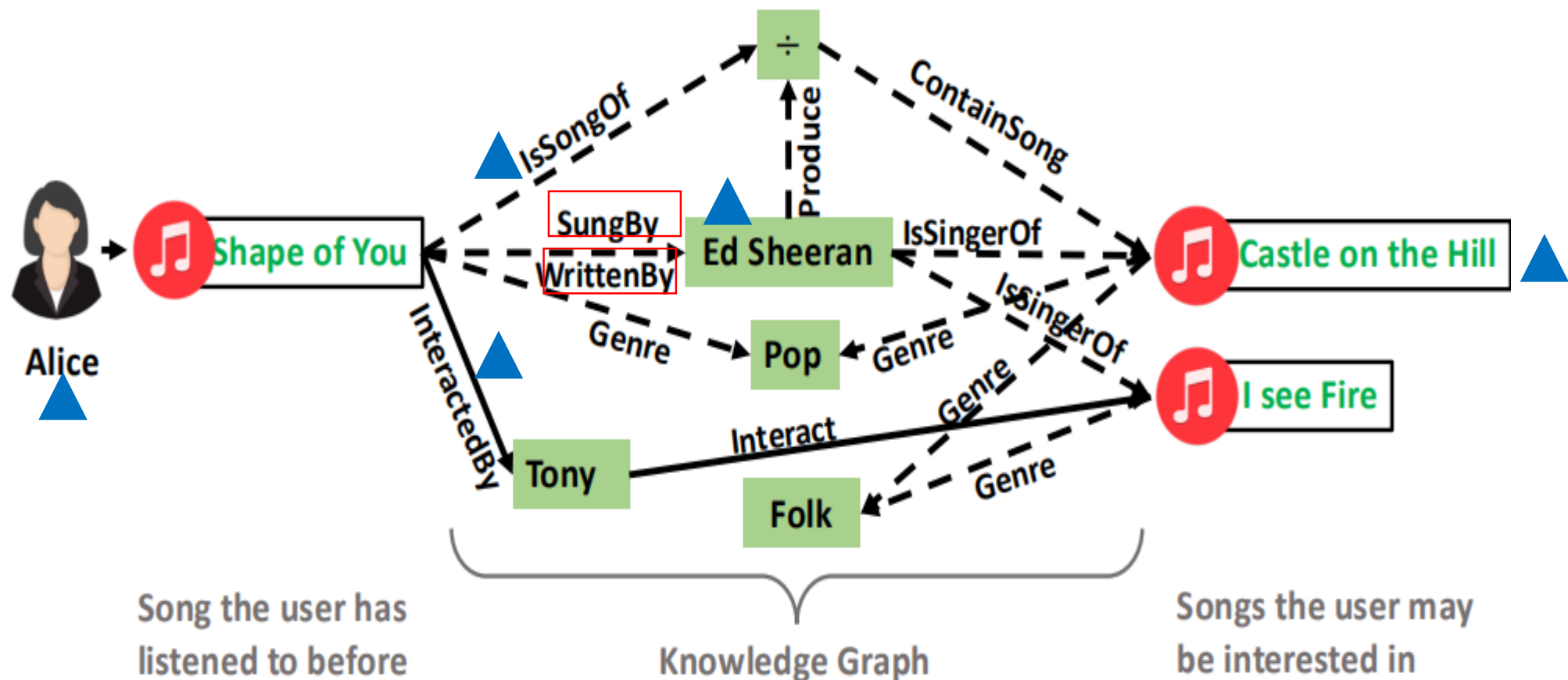
Knowledge-aware Path Recurrent Network

常见符号

we define KG as $\mathcal{KG} = \{(h, r, t) | h, r \in \mathcal{E}, r \in \mathcal{R}\}$, where each triplet (h, r, t) indicates a fact that there is a relationship r from head entity h to tail entity t .

The user-item interaction data is usually presented as a bipartite graph. In particular, we use $\mathcal{U} = \{u_t\}_{t=1}^M$ and $\mathcal{I} = \{i_t\}_{t=1}^N$ to separately denote the user set and the item set, where M and N are the number of users and items,

we represent the interaction between a user and an item with a triplet $\tau = (u, \text{interact}, i)$, if there is an observed interaction (e.g., rate, click, and view feedbacks), where interact is a pre-defined relation.



- 用户和他们可能的交互背后的项目之间的复杂关系（即路径）激发了通过考虑**两个实体和（多步骤）组合地模拟路径的高级语义关系**。
- 因此，从推理的角度来看，我们**消耗所有路径上的连通性来学习组合关系表示**，并将它们**加权汇集在一起**以预测用户和目标项之间的交互关系。

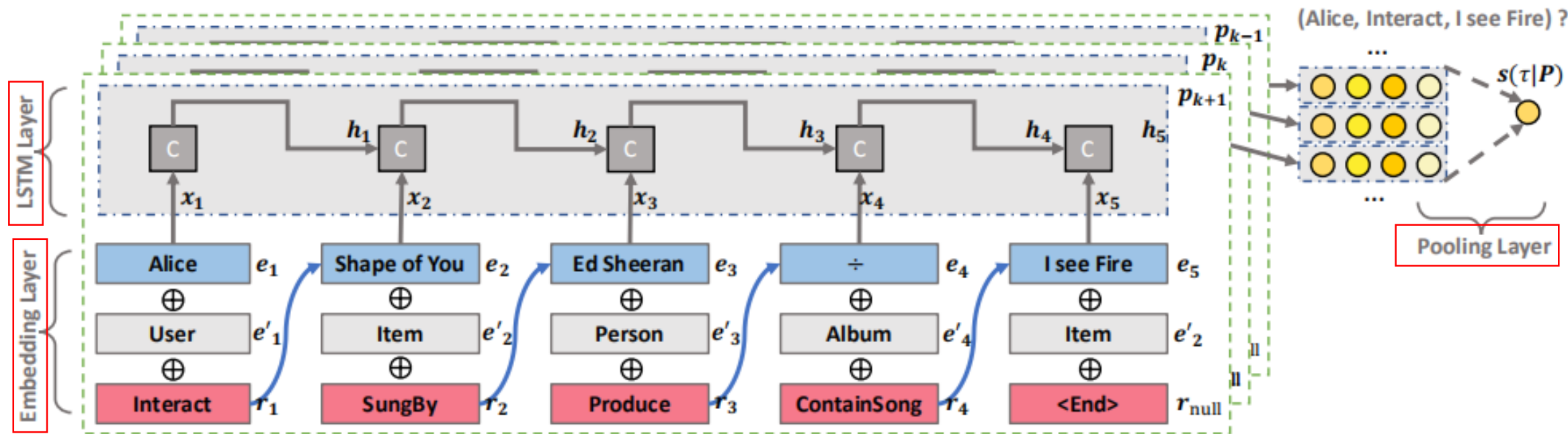
目标函数

给定用户 u ，目标物品 i ，和连接两者的路径集合 \mathcal{P} ，则整体目标函数定义为：

$$\hat{y}_{ui} = f_{\Theta}(u, i | \mathcal{P}(u, i)), \quad (1)$$

where f denotes the underlying model with parameters Θ , and \hat{y}_{ui} presents the predicted score for the user-item interaction. Distinct from embedding-based methods, we can explain \hat{y}_{ui} as the plausibility score of the triplet $\tau = (u, \text{interact}, i)$ inferred by the connectivity $\mathcal{P}(u, i)$.

建立模型



- Embedding Layer: 针对路径 p 得到嵌入的一个集合，其中每一个元素表示实体或关系

- LSTM Layer

$$\mathbf{x}_{l-1} = \mathbf{e}_{l-1} \oplus \mathbf{e}'_{l-1} \oplus \mathbf{r}_{l-1},$$

$$\mathbf{z}_l = \tanh(\mathbf{W}_z \mathbf{x}_l + \mathbf{W}_h \mathbf{h}_{l-1} + \mathbf{b}_z)$$

$$\mathbf{f}_l = \sigma(\mathbf{W}_f \mathbf{x}_l + \mathbf{W}_h \mathbf{h}_{l-1} + \mathbf{b}_f)$$

$$\mathbf{i}_l = \sigma(\mathbf{W}_i \mathbf{x}_l + \mathbf{W}_h \mathbf{h}_{l-1} + \mathbf{b}_i)$$

$$\mathbf{o}_l = \sigma(\mathbf{W}_o \mathbf{x}_l + \mathbf{W}_h \mathbf{h}_{l-1} + \mathbf{b}_o)$$

$$\mathbf{c}_l = \mathbf{f}_l \odot \mathbf{c}_{l-1} + \mathbf{i}_l \odot \mathbf{z}_l$$

$$\mathbf{h}_l = \mathbf{o}_l \odot \tanh(\mathbf{c}_l)$$

$$s(\tau|\mathbf{p}_k) = \mathbf{W}_2^\top \text{ReLU}(\mathbf{W}_1^\top \mathbf{p}_k),$$

- Weighted Pooling Layer

$$\hat{y}_{ui} = \sigma\left(\frac{1}{K} \sum_{k=1}^K s_k\right).$$

$$g(s_1, s_2, \dots, s_K) = \log \left[\sum_{k=1}^K \exp\left(\frac{s_k}{\gamma}\right) \right]$$

and the final prediction score is given by,

$$\hat{y}_{ui} = \sigma(g(s_1, s_2, \dots, s_K)), \quad (7)$$

where γ is the hyper-parameter to control each exponential weight. Such pooling is capable of distinguishing the path importance, which is attributed by the gradient:

$$\frac{\partial g}{\partial s_k} = \frac{\exp(s_k/\gamma)}{\gamma \sum_{k'} \exp(s_{k'}/\gamma)}, \quad (8)$$

参数学习

In particular, the negative log-likelihood is adopted as the objective function, which is defined as follows,

$$\mathcal{L} = - \sum_{(u,i) \in \mathcal{O}^+} \log \hat{y}_{ui} + \sum_{(u,j) \in \mathcal{O}^-} \log(1 - \hat{y}_{uj}), \quad (9)$$

where $\mathcal{O}^+ = \{(u,i) | y_{ui} = 1\}$ and $\mathcal{O}^- = \{(u,j) | y_{uj} = 0\}$ are the positive and negative user-item interaction pairs,

实验

● 数据集

Table 1: Statistics of our datasets.

	Dataset	MI	KKBox
User-Item Interaction	#Users	6,040	34,403
	#Items	3,859	2,296,833
	#Interactions	998,034	3,714,655
Knowledge Graph	#Entities	11,462	2,851,220
	#Entity Types	4	4
	#Relation Types	6	6
	#Triplets	1,017,030	11,182,682
Path	#Paths	55,573,556	38,192,484
	Avg Path Length	5.07	5.09

● 路径提取

提取连接所有用户-物品对的所有限定路径，每个路径长度最多为6。

● 实验设置

● 评价指标：

Hit@K

Ndcg@k

● 基本方法

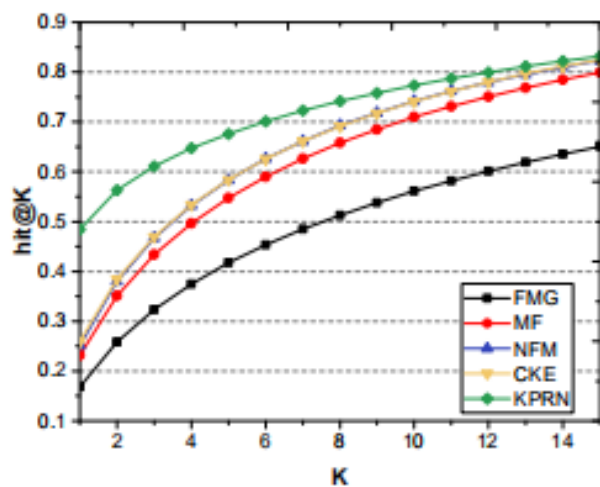
MF

NeuFM：将神经网络应用到因子分解机中。

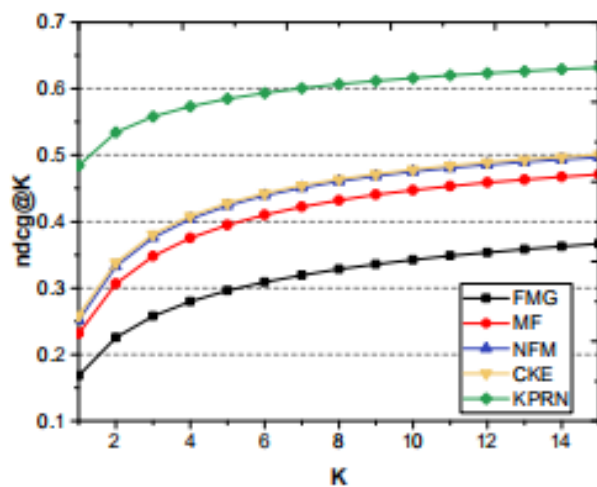
CKE：一种利用知识图谱嵌入的推荐模型，集成了Matrix Factorization和TransR来提高推荐效果。

FMG：这是一种先进的基于元路径的方法，它预定义了各种类型的元图，并在每个元图相似矩阵上采用矩阵分解来进行推荐。

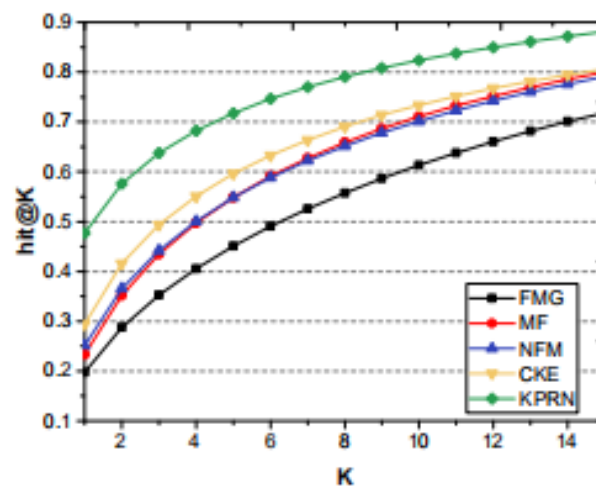
● 性能对比



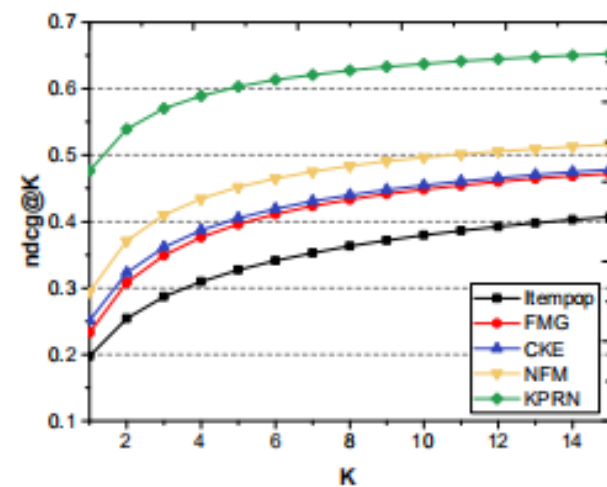
(a) $\text{hit}@K$ on MI



(b) $\text{ndcg}@K$ on MI



(c) $\text{hit}@K$ on KKBox



(d) $\text{ndcg}@K$ on KKBox

Figure 3: Top- K recommendation performance between all the methods on MI and KKBox datasets *w.r.t.* $\text{hit}@K$ and $\text{ndcg}@K$, where $K = \{1, 2, \dots, 15\}$.

- FMG方法效果不好
- NFM方法比FM表现要好
- CKE方法对稀疏数据的效果较好
- KPRN性能最佳

● KPRN方法

首先探讨路径中关系的影响。然后，研究加权池操作如何影响性能。

关系建模的影响

没有关系建模的KPRN变体，称为KPRN-r。

具体地，丢弃等式（2）中的关系嵌入以生成输入向量。

◆ KPRN-r的性能在两个数据集上都会降低。

◆ 我们发现在MI数据集中KPRN将KPRN-r提高了6.45%，而在KKBox上只有0.70%。

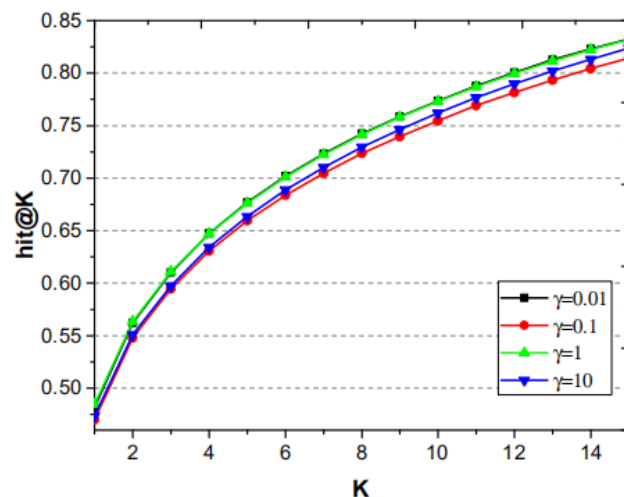
Table 2: Performance comparison of KPRN and KPRN-r and their effects on relation modeling.

	MI						KKBox					
	hit@5	hit@10	hit@15	ndcg@5	ndcg@10	ndcg@15	hit@5	hit@10	hit@15	ndcg@5	ndcg@10	ndcg@15
KPRN-r	0.635	0.738	0.801	0.533	0.566	0.583	0.712	0.821	0.878	0.607	0.632	0.647
KPRN	0.676	0.773	0.832	0.584	0.616	0.632	0.717	0.823	0.881	0.613	0.637	0.652

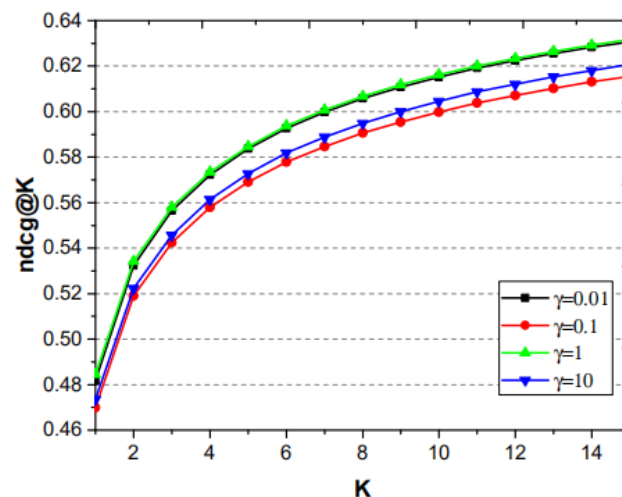
加权池的影响

为了整合用户-项对之间多路径的预测分数，仔细设计加权池操作。为了分析其影响，我们将值 γ 设置为 $\{0.01, 0.1, 1, 10\}$ 。

- ◆ 当 γ 从1减小到0.1时，加权池操作会降低性能，因为它类似于max-pooling，只选择最重要的路径作为用户项连接。
- ◆ 当将 γ 从1增加到10时，Hit@ K和ndcg @ K变得更差。



(a) hit@K on MI



(b) ndcg@K on MI

Figure 4: Performance comparison of KPRN *w.r.t.* γ on the MI dataset.

● 案例分析

KPRN的另一个理想特性是**在路径上推断用户对目标物品的偏好**并产生合理的解释。

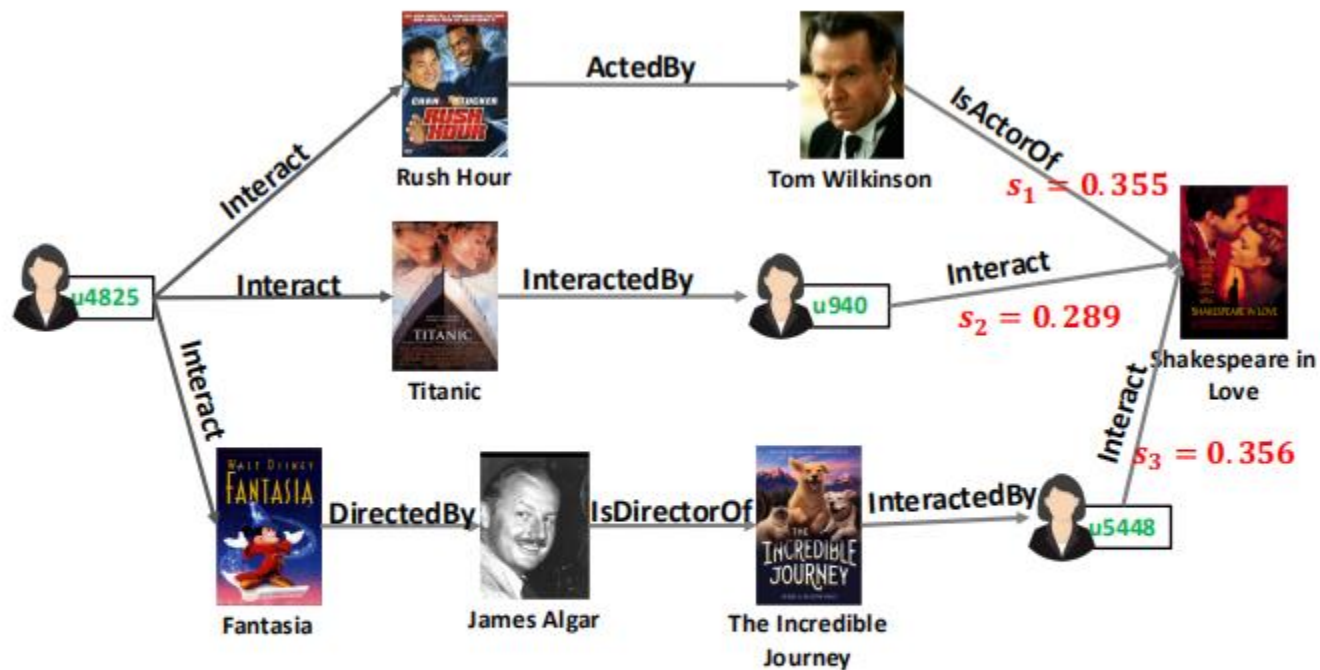


Figure 5: Visualization of three paths with prediction scores for the user of u_{4825} in MI dataset. The prediction scores are normalized for illustration.

谢 谢