RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Neural-Symbolic Reasoning over Knowledge Graph for Multi-stage Explainable Recommendation

Yikun Xian, Zuohui Fu, Qiaoying Huang
S. Muthukrishnan, Yongfeng Zhang

Department of Computer Science
Rutgers University

Rutgers
THE STATE UNIVERSITY OF NEW JERSEY

Outline

- Introduction
- Methodology
- Experiment Results
- Conclusion

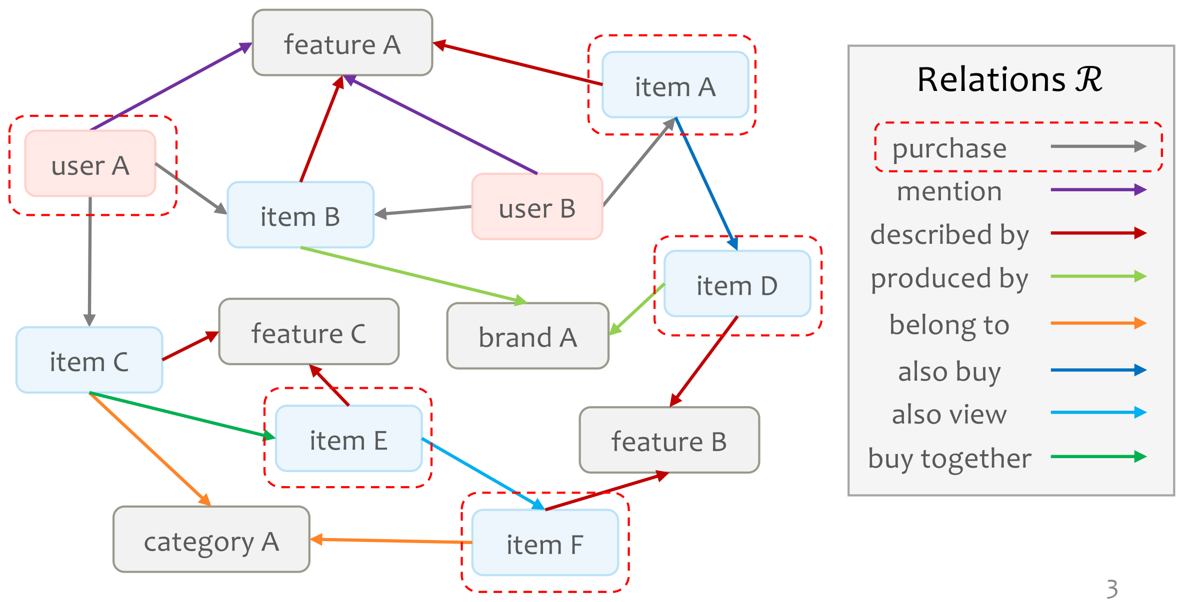
KG-based Explainable Recommendation

Knowledge Graph (KG)

- A knowledge graph $\mathcal{G}$ with entity set $\mathcal{E}$ and relation set $\mathcal{R}$ is defined as $\mathcal{G} = \{(e_h, r, e_t) | e_h, e_t \in \mathcal{E}, r \in \mathcal{R}\}$, where each triplet (e_h, r, e_t) represents a fact of the relation r from head entity e_h to tail entity e_t.

KG for Recommendation

+ User entity set $\mathcal{U} \subset \mathcal{E}$.
+ Item entity set $\mathcal{I} \subset \mathcal{E}$.
+ Relation "purchase".



KG-based Explainable Recommendation

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

<u>Problem Definition</u>

- **Inputs**: knowledge graph $\mathcal{G}$, user $u \in \mathcal{U}$ and integer K, N.
- **Outputs**: a set of recommended items $\{i_n\}_{n\in[N]} \subseteq \mathcal{I}$ such that each pair (u, i_n) is associated with one reasoning path $p_k(u, i_n)$, $2 \leq k \leq K$.
  - N: number of recommendations for each user.
  - $p_k(u, i_n)$: a path of length k from node u to node i_n.

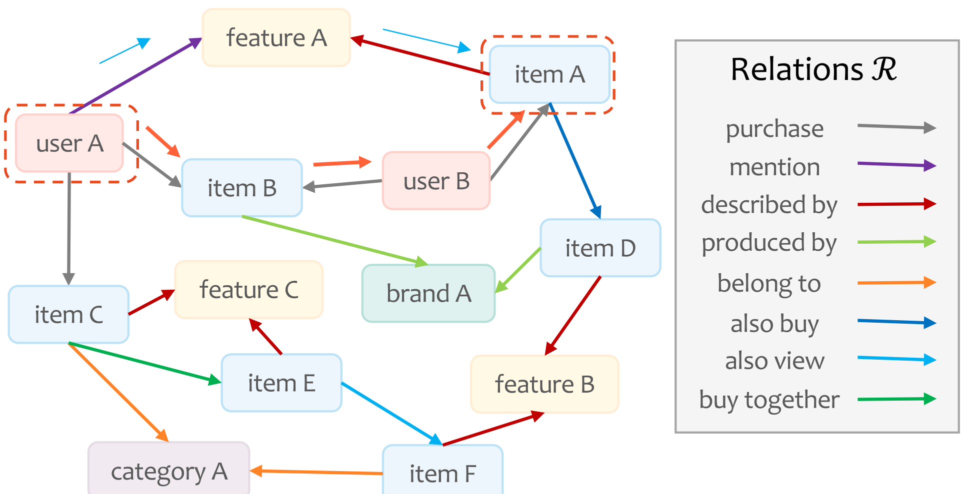
<u>Challenges</u>

- Unknown target: items (target node) are NOT known before path finding.
- Large node degree: this leads to large search space.

KG-based Explainable Recommendation

Example

Inputs: KG, u=“user A”, K=3, N=3.
(Assume “item A” is potential recommendation.)



Outputs:
One of the following paths:

- “user A” → “purchase” → “item B” → “purchase” → “user B” → “purchase” → “item A”
- “user A” → “mention” → “feature A” → “described_by” → ”item A”

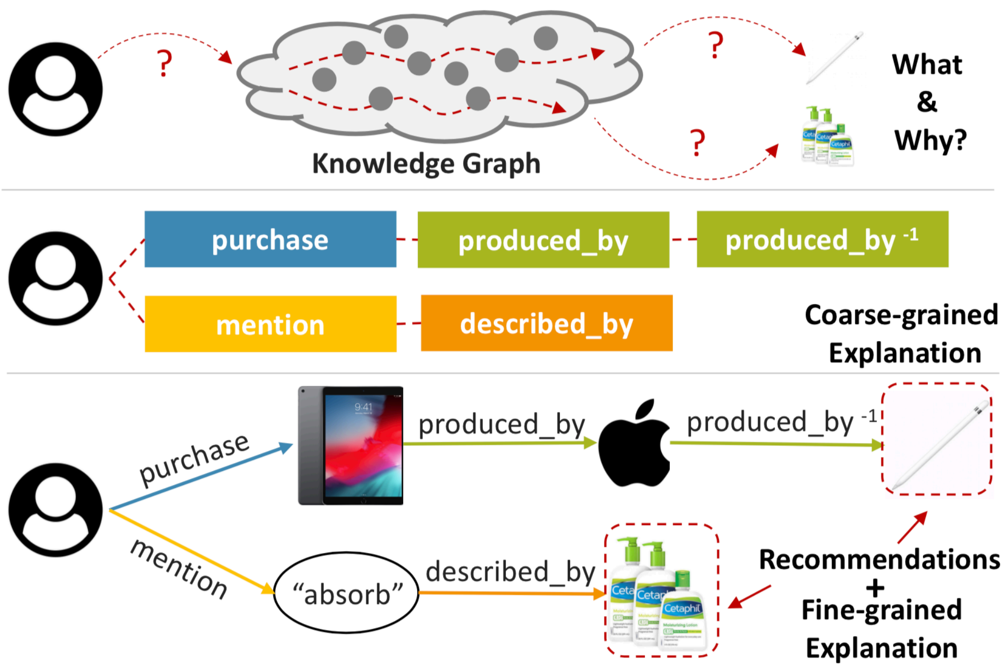
Related Work

- Pre-defined & Post-hoc
  - Most existing works generate paths in either the *pre-defined* or the *post-hoc* manner.
  - However, the resulting paths cannot reflect the decision-making process of the models.

- In-progress
  - Limited works generate paths in the *in-progress* manner.
  - However, the recommendation performance is weakened by the path reasoning due to large search space over KG.

Motivation

- Can we first generate a coarse-grained explanation to portray "user profile"?
- Then perform path reasoning over KG based on the "profile"?



RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Outline

- Introduction ✓
- Methodology ☛
- Experiment Results
- Conclusion

A Neural-Symbolic Approach

- Goal
  - Given a user over a KG, how to find a set of paths that eventually arrives at some items of interest.
  - Explicit path reasoning over KG.
  - Good recommendation performance.

- Our Method
  - Neural-Symbolic Representation Learning
  - Neural-Symbolic Explainable Recommendation

Neural-Symbolic Representation Learning

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

- Question 1

  How to empower the model with path reasoning ability?

- Answer

  Learning from training paths by maximizing the log likelihood of the paths being generated by our model.

Neural-Symbolic Representation Learning

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

- Given a user–item path $L = \{u, r_1, e_1, r_2, e_2, \ldots, e_{|L|-1}, r_{|L|}, i\}$,
- Maximize log likelihood of $P(L|u;\Theta)$ with model parameters Θ:

$$\log P(L|u;\Theta) = \sum_{t=1..|L|} \log P(e_t, r_t | u, h_t; \Theta)$$

$$P(e_t, r_t | u, h_t; \Theta) = \frac{1}{Z} \exp(s(e_t, r_t, u, h_t; \Theta))$$

where

- $h_t = \{u, r_1, e_1, \ldots, r_{t-1}, e_{t-1}\}$ is the history prior to the t-th step.
- Z is a normalization term.
- $s(e_t, r_t, u, h_t; \Theta)$ is a score indicating how likely node e_t is being generated via relation r_t given u, h_t.

Neural-Symbolic Representation Learning

- We estimate $s(e_t, r_t, u, h_t; \Theta)$ via a neural relation module ϕ_{r_t}:

$$s(e_t, r_t, u, h_t; \Theta) = \langle \boldsymbol{e}_t, \phi_{r_t}(\boldsymbol{u}, \boldsymbol{h}_t; \Theta) \rangle$$
$$\phi_{r_t}(\boldsymbol{u}, \boldsymbol{h}_t; \Theta) = W_3 \sigma(W_2 \sigma(W_1 [\boldsymbol{u}; \boldsymbol{h}_t]))$$

$\boldsymbol{u}$ is the input user vector to control personalization.

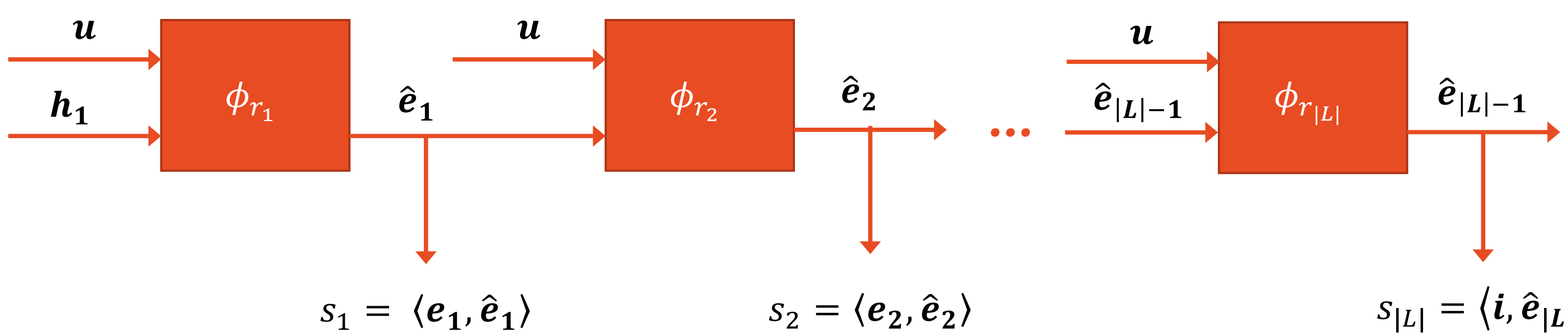
$\boldsymbol{h}_t$ is the output from the previous module $\phi_{r_{t-1}}$.

Output vector $\hat{\boldsymbol{e}}_t$ is used to find the possible next-hop entities!

- For each relation r in the knowledge graph, we maintain a neural relation module ϕ_r.

Neural-Symbolic Representation Learning

- $L = \{u, r_1, e_1, r_2, e_2, \dots, e_{|L|-1}, r_{|L|}, i\}$



$$\log P(L|u;\Theta) = \log \exp(s_1) - \log Z_1 + \log \exp(s_2) - \log Z_2 + \cdots + \log \exp\left(s_{|L|}\right) - \log Z_{|L|}$$

- Model parameters Θ include entity embedding + neural relation modules.

Neural-Symbolic Representation Learning

- For all users $\{u\}$ along with the training paths set $\{L_u\}$,
- We can estimate model parameters $\widehat{\Theta}$:

$$\widehat{\Theta} = \operatorname{argmax}_{\Theta} \sum_{u} \ell_{path}(\Theta; \{L_u\})$$

$$\underline{\ell_{path}(\Theta; \{L_u\})} = \sum_{L_u} \log P(L_u | u; \Theta)$$

Path regularization: composition of neural relation modules
is more likely to generate paths that assemble $\{L_u\}$.

Neural-Symbolic Representation Learning

- Question 2

  How to guarantee recommendation performance?

- Answer

  Imposing an additional ranking loss to distinguish the quality of reached items.

Neural-Symbolic Representation Learning

- For each user u along with the path set $\{L_{\mathrm{u}}\}$, and
- For each path L_u, there is a set of negative items $\{i^-\}$ w.r.t (u, i^+),
- The goal is to minimize the ranking loss ℓ_{rank}:

$$\ell_{rank}(\Theta; \{L_u\}) = \sum_{L_u} \sum_{i^-} \sigma\left(s\left(i^-, r_{|L_u|}, u, h_{|L_u|}\right) - s\left(i^+, r_{|L_u|}, u, h_{|L_u|}\right)\right)$$

$$s\left(i, r_{|L_u|}, u, h_{|L_u|}\right) = \left\langle \boldsymbol{i}, \phi_{r_{|L_u|}}(\boldsymbol{u}, \boldsymbol{h_t}; \Theta) \right\rangle$$

Neural-Symbolic Representation Learning

- For each user u along with the path set $\{L_{\mathrm{u}}\}$, and
- A set of negative items $\{i^-\}$ w.r.t each path $L_{\mathrm{u}}: u \rightsquigarrow i^+$,
- The goal is to minimize the ranking loss ℓ_{rank}:

$$\ell_{rank}(\Theta; \{L_u\}) = \sum_{L_u} \sum_{i^-} \sigma\left(s\left(i^-, r_{|L_u|}, u, h_{|L_u|}\right) - s\left(i^+, r_{|L_u|}, u, h_{|L_u|}\right)\right)$$

$$s\left(i, r_{|L_u|}, u, h_{|L_u|}\right) = \left\langle \boldsymbol{i}, \phi_{r_{|L_u|}}(\boldsymbol{u}, \boldsymbol{h_t}; \Theta)\right\rangle$$

Pairwise ranking: the model is more likely to reach a positive item i^+ than the negative items $\{i^-\}$.

Neural-Symbolic Representation Learning

- Overall objective

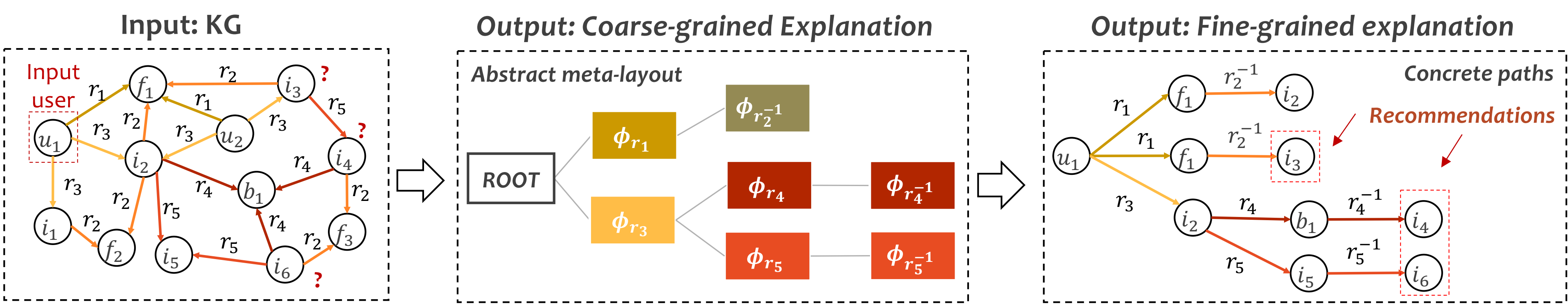
$$\ell_{all}(\Theta) = \sum_{u} \ell_{path}(\Theta; \{L_u\}) + \lambda \ell_{rank}(\Theta; \{L_u\})$$

where λ is weighting factor over ranking loss.

Neural-Symbolic Explainable Recommendation

A two-stage approach

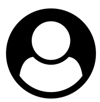
- Coarse-grained explanation: construct a personalized layout of neural network structure.
- Fine-grained explanation: perform path reasoning with the composed network for recommendation.



Neural-Symbolic Explainable Recommendation

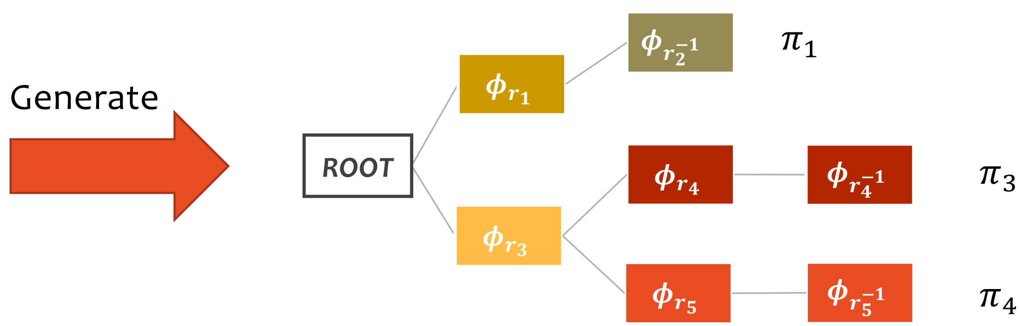
Generating Coarse-grained Explanation (Layout)

- Given a user u and a set of patterns (rules) $\{\pi_1, \ldots, \pi_M\}$, the goal is to construct a tree-structured layout such that the selected patterns can be used to generate reasoning paths with high probability.



u

$$\begin{aligned}
\pi_1 &= \{r_1, r_2^{-1}\} \\
\pi_2 &= \{r_1, r_1^{-1}, r_3\} \\
\pi_3 &= \{r_3, r_4, r_4^{-1}\} \\
\pi_4 &= \{r_3, r_5, r_5^{-1}\} \\
\pi_5 &= \{r_3, r_2, r_2^{-1}\} \\
&\ldots
\end{aligned}$$



Abstract meta-layout

Neural-Symbolic Explainable Recommendation

Generating Coarse-grained Explanation (Layout)

- Solution:
  - Given a user u, for each pattern π, we estimate a heuristic score

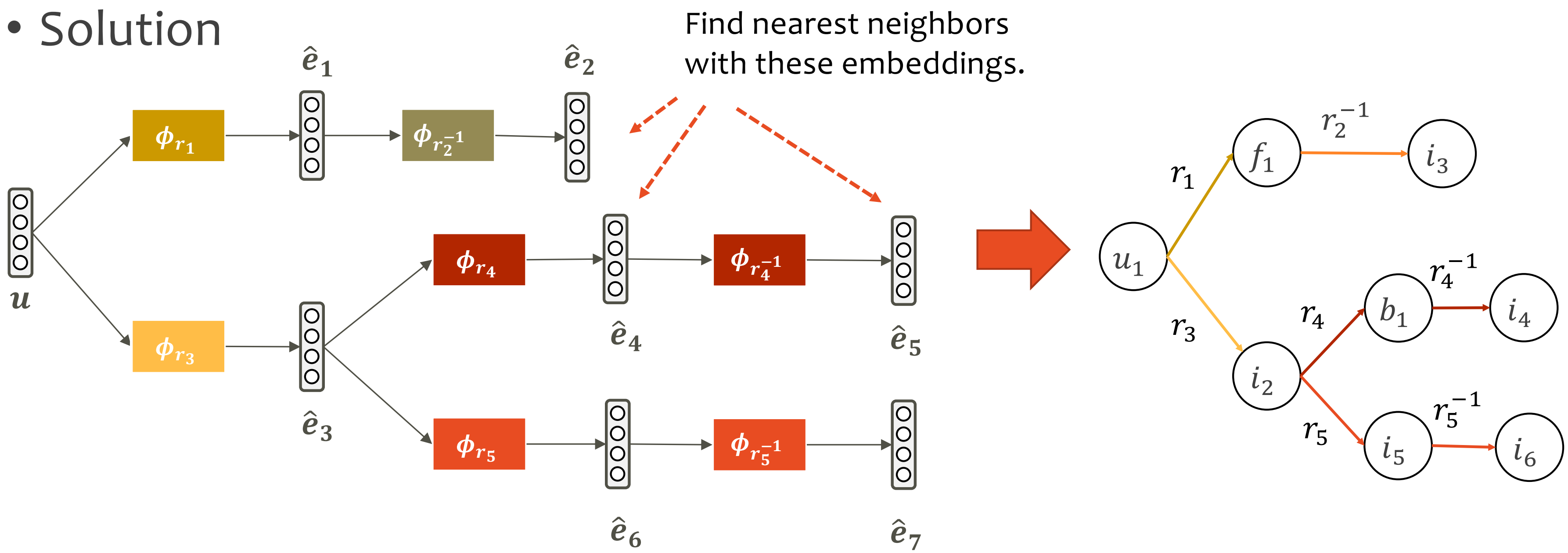
$$v(u,\pi) = E_{L_u^\pi}[\log P(L_u^\pi | u; \Theta)]$$

  where
  - L_u^π is the path of pattern π starting from u, randomly sampled from training set.
  - For each user, select top patterns $\{\pi\}$ with the largest score $v(u,\pi)$.
  - Aggregate all the selected patterns $\{\pi\}$.

Neural-Symbolic Explainable Recommendation

Generating Fine-grained Explanation (Paths)

- Given the user and the layout, how to generate a set of paths for recommendation?
- Solution



Outline

- Introduction ✓
- Methodology ✓
- Experiment Results
- Conclusion

Datasets

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

- 4 Amazon e-commerce datasets

		CDs & Vinyl	Clothing	Cell Phones	Beauty
Entities	**Description**	**Number of Entities**			
User	User in recommender system.	75,258	39,387	27,879	22,363
Item	Product to be recommended to users.	64,443	23,033	10,429	12,101
Feature	A product feature word from reviews.	202,959	21,366	22,493	22,564
Brand	Brand or manufacturer of the product.	1,414	1,182	955	2,077
Category	Category of the product.	770	1,193	206	248
Relations	**Description**	**Number of Relations per Head Entity**			
Purchase	$User \xrightarrow{\text{purchase}} Item$	14.58 ± 39.13	7.08 ± 3.59	6.97 ± 4.55	8.88 ± 8.16
Mention	$User \xrightarrow{\text{mention}} Feature$	$2,545.92 \pm 10,942.31$	440.20 ± 452.38	652.08 ± 1335.76	806.89 ± 1344.08
Described_by	$Item \xrightarrow{\text{described\_by}} Feature$	$2,973.19 \pm 5,490.93$	752.75 ± 909.42	$1,743.16 \pm 3,482.76$	$1,491.16 \pm 2,553.93$
Belong_to	$Item \xrightarrow{\text{belong\_to}} Category$	7.25 ± 3.13	6.72 ± 2.15	3.49 ± 1.08	4.11 ± 0.70
Produced_by	$Item \xrightarrow{\text{produced\_by}} Brand$	0.21 ± 0.41	0.17 ± 0.38	0.52 ± 0.50	0.83 ± 0.38
Also_bought	$Item \xrightarrow{\text{also\_bought}} Item$	57.28 ± 39.22	61.35 ± 32.99	56.53 ± 35.82	73.65 ± 30.69
Also_viewed	$Item \xrightarrow{\text{also\_viewed}} \text{another } Item$	0.27 ± 1.86	6.29 ± 6.17	1.24 ± 4.29	12.84 ± 8.97
Bought_together	$Item \xrightarrow{\text{bought\_together}} \text{another } Item$	0.68 ± 0.80	0.69 ± 0.90	0.81 ± 0.77	0.75 ± 0.72

Table 1: Descriptions and statistics of four Amazon e-commerce datasets: *CDs & Vinyl*, *Clothing*, *Cell Phones* and *Beauty*.

Main Results

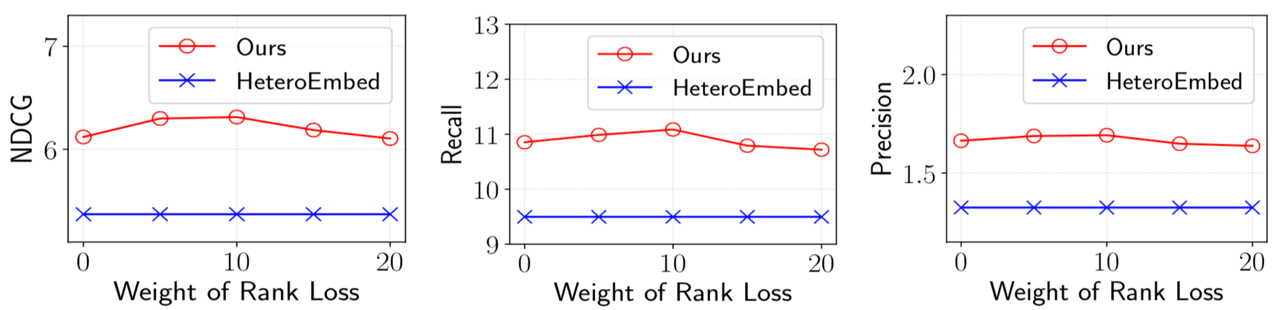
RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

- Recommendation performance
  - Our method outperforms all other baselines on all datasets in terms of NDCG, Hit Rate, Recall and Precision.

Dataset	**CDs & Vinyl**				**Clothing**				**Cell Phones**				**Beauty**			
Measures (%)	NDCG	Recall	HR	Prec.	NDCG	Recall	HR	Prec.	NDCG	Recall	HR	Prec.	NDCG	Recall	HR	Prec.
DeepCoNN	4.218	6.001	13.857	1.681	1.310	2.332	3.286	0.229	3.636	6.353	9.913	0.999	3.359	5.429	9.807	1.200
CKE	4.620	6.483	14.541	1.779	1.502	2.509	4.275	0.388	3.995	7.005	10.809	1.070	3.717	5.938	11.043	1.371
HeteroEmbed	5.563	7.949	17.556	2.192	3.091	5.466	7.972	0.763	5.370	9.498	13.455	1.325	6.399	10.411	17.498	1.986
PGPR	5.590	7.569	16.886	2.157	2.858	4.834	7.020	0.728	5.042	8.416	11.904	1.274	5.449	8.324	14.401	1.707
NSER (Ours)	**6.868**	**9.376**	**19.692**	**2.562**	**3.689**	**6.340**	**9.275**	**0.975**	**6.313**	**11.086**	**15.531**	**1.692**	**7.061**	**10.948**	**18.099**	**2.270**

Influence of Ranking Loss

- When no ranking loss is imposed, our method still outperforms the baseline.
- The best performance is achieved when the weight is around 10.



(a) NDCG (b) Recall (c) Precision

Figure 3: Ranking loss results on Cell Phones dataset.

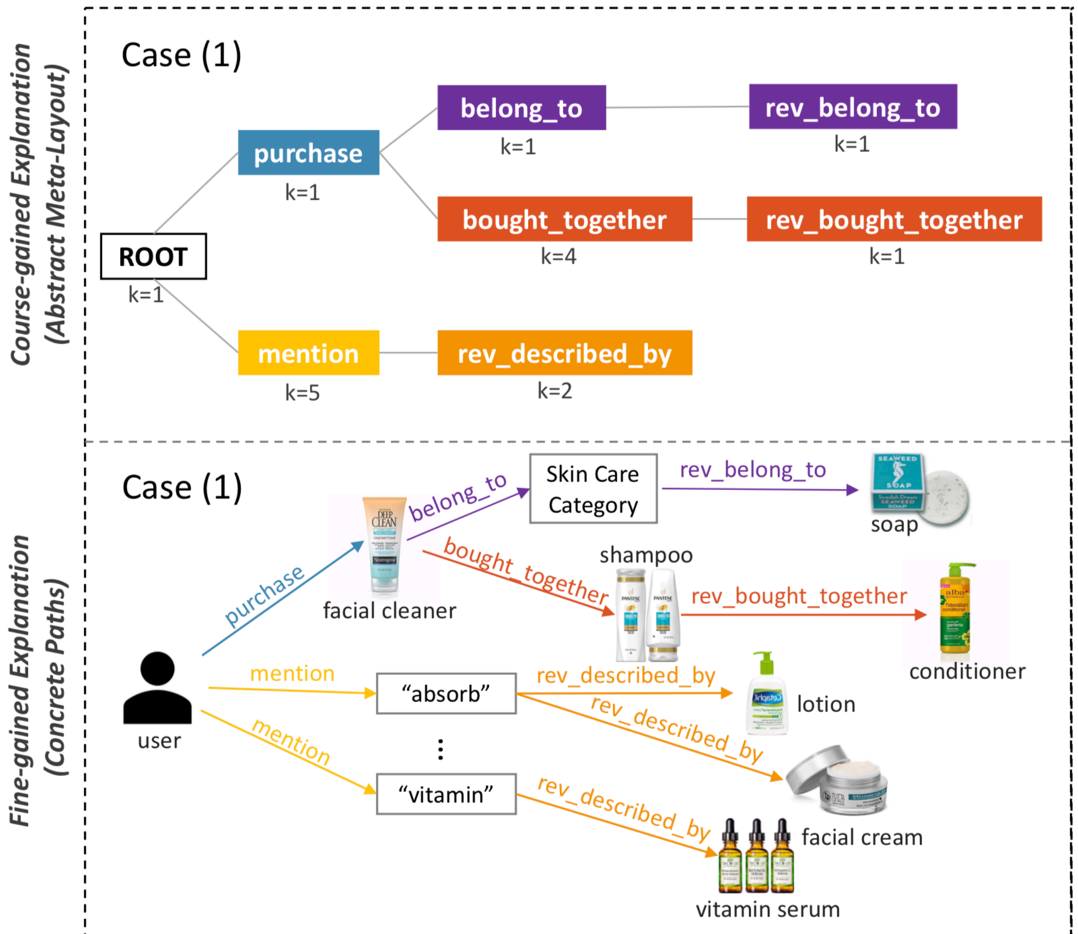
Effectiveness of Layout

- Uniform: randomly choose a set of patterns to construct the layout.
- Prior: choose patterns based on frequency of training paths.
- Heuristic: the proposed method.

Dataset	**Cell Phones**				**Beauty**			
Method	NDCG	Recall	HR	Prec.	NDCG	Recall	HR	Prec.
uniform	4.545	7.229	10.192	1.087	6.293	9.256	15.564	1.918
prior	6.255	10.842	15.097	1.659	6.880	10.393	17.258	2.224
heuristic	6.313	11.086	15.531	1.692	7.061	10.948	18.099	2.270

Case Study

- Example



Outline

- Introduction ✓
- Methodology ✓
- Experiment Results ✓
- Conclusion

Conclusion

- A neural-symbolic reasoning approach for explainable recommendation over KG
- Generate explanations in two stages:
  - A coarse-grained explanation (abstract layout)
  - A find-grained explanation (concrete paths)
- Experimental results show promising recommendation performance by our method.

RUTGERS

Q & A