Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preference

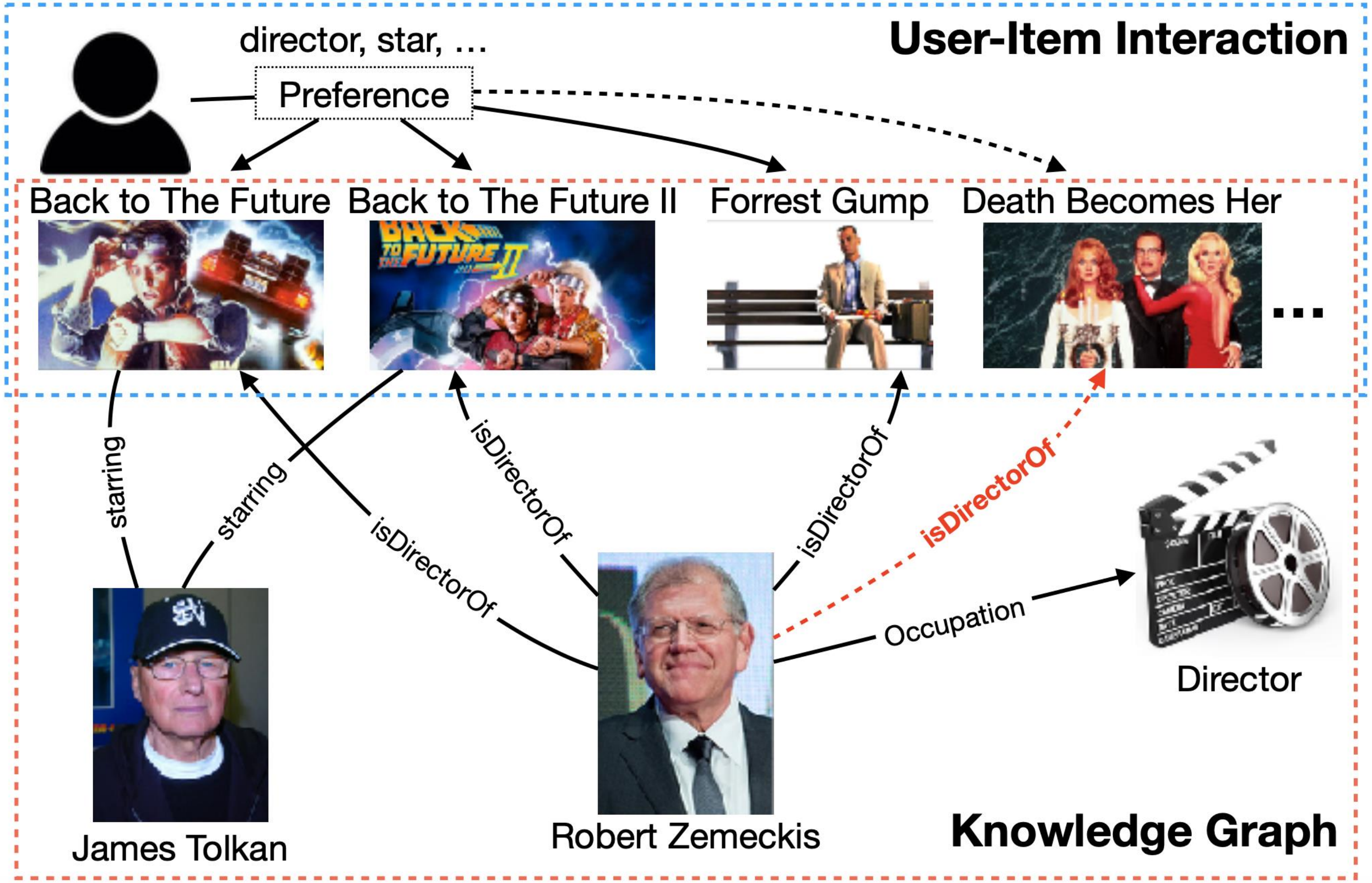
서울시립대학교
UNIVERSITY OF SEOUL

WWW'19

cheon.research@gmail.com

2020. 03. 31.

KG for Recommendation



서울시립대학교 UNIVERSITY OF SEOUL

Tasks and Notations

Item Recommendation : Given a list of user-item interactions $\mathcal{Y} = \{(u, i)\}$, we use implicit feedback as the protocol so that each pair (u, i) implies the user $u \in \mathcal{U}$ consumes the item $i \in \mathcal{I}$. The goal is to recommend top-N items for a target user.

KG Completion : A **Knowledge Graph** $\mathcal{KG}$ is a directed graph composed of *subject-property-object* triple facts. Each triplet denotes that there is a relationship r from head entity e_h to tail entity e_t, formally defined by (e_h, e_t, r), where $e_h, e_t \in \mathcal{E}$ are entities and $r \in \mathcal{R}$ are relations. Due to the incompleteness nature of KGs, KG completion is to predict the missing entity e_h or e_t for a triplet (e_h, e_t, r), which can also be regarded as recommending top-N entities for a target (e_t, r) or (e_h, r).

Tasks and Notations

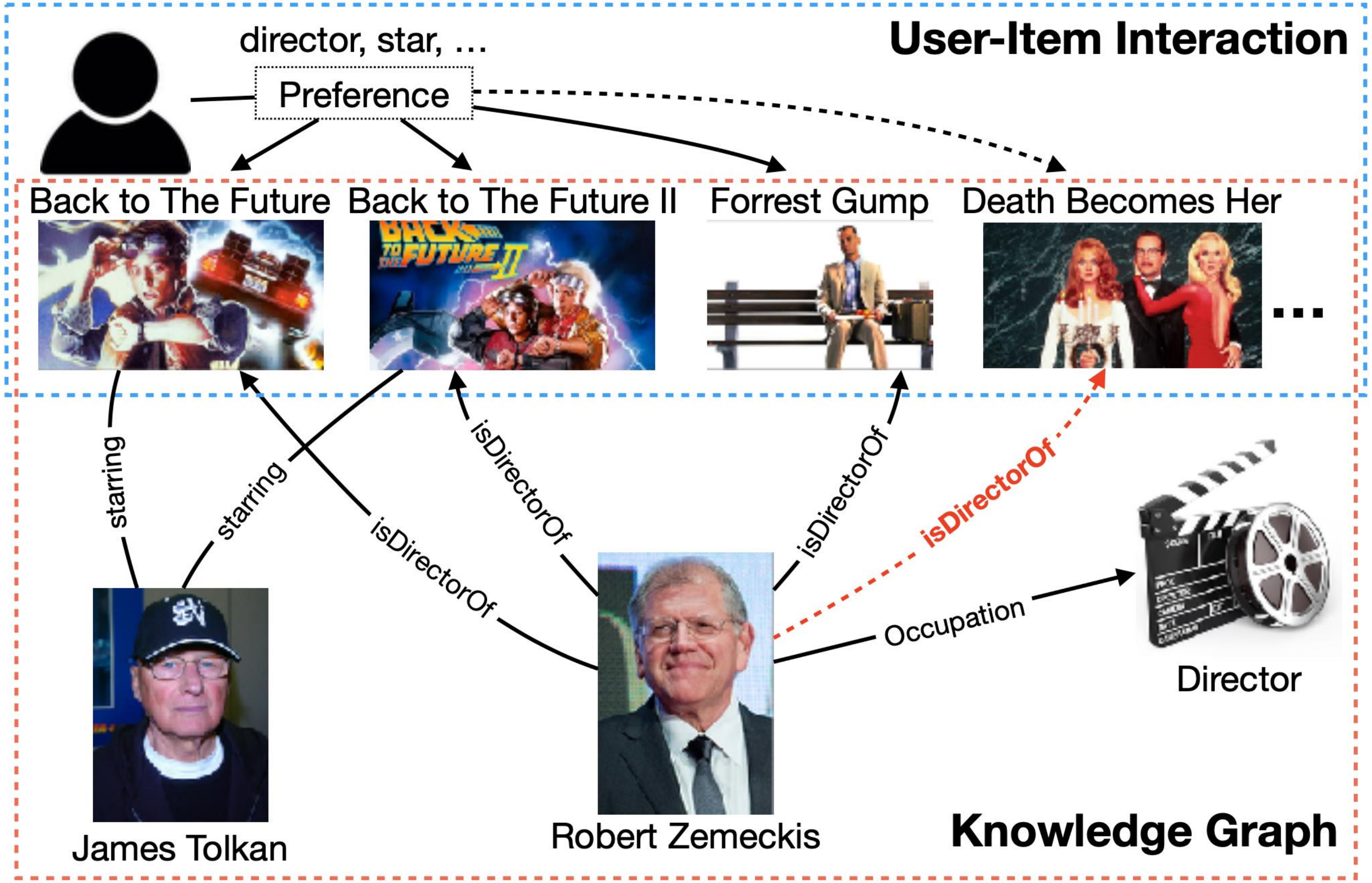
TUP denotes the model for item recommendation. It takes a list of user-item pairs $\mathcal{Y}$ as input, and outputs a relevance score $g(u, i; p)$ indicating the likelihood that u likes i, given the preference $p \in \mathcal{P}$, where the number of the preference set $\mathcal{P}$ is predefined. For each user-item pair, we induce a preference, serving as a similar role with the relation for two entities. To deal with the N-to-N issue, we introduce preference hyperplanes, and assign each preference with two vectors: $\mathbf{w}_p$ for the projection to a hyperplane, $\mathbf{p}$ for the translation between users and items.

서울시립대학교
UNIVERSITY OF SEOUL

Tasks and Notations

KTUP is a multi-task architecture. Given $\mathcal{KG}$, $\mathcal{Y}$, and a set of item-entity alignments $\mathcal{A} = \{(i, e)|i \in \mathcal{I}, e \in \mathcal{E}\}$, where each (i, e) means that i can be mapped to an entity e in the given KG. KTUP is able to output not only $g(u, i; p)$, but also a score $f(e_h, e_t, r)$ indicating how possible the fact is true, based on the jointly learned embeddings of users $\mathbf{u}$, items $\mathbf{i}$, preferences $\mathbf{p}, \mathbf{w}_p$, entities $\mathbf{e}$ and relations $\mathbf{r}, \mathbf{w}_r$.

KG for Recommendation



item-entity alignments

TransH for KG Completion

$$\text{triplet } (e_h, e_t, r)$$

$$f(e_h, e_t, r) = \| \mathbf{e}_h^{\perp} + \mathbf{r} - \mathbf{e}_t^{\perp} \| \tag{1}$$

where a lower score of $f(e_h, e_t, r)$ indicates that the triplet is poss-bily true, otherwise no. $\mathbf{e}_h^{\perp}$ and $\mathbf{e}_t^{\perp}$ are projected entity vectors:

$$\mathbf{e}_h^{\perp} = \mathbf{e}_h - \mathbf{w}_r^{\mathrm{T}} \mathbf{e}_h \mathbf{w}_r \tag{2}$$

$$\mathbf{e}_t^{\perp} = \mathbf{e}_t - \mathbf{w}_r^{\mathrm{T}} \mathbf{e}_t \mathbf{w}_r \tag{3}$$

TransH for KG Completion

$$\mathcal{L}_k = \sum_{(e_h, e_t, r) \in \mathcal{KG}} \sum_{(e'_h, e'_t, r') \in \mathcal{KG}^-} [f(e_h, e_t, r) + \gamma - f(e'_h, e'_t, r')]_+ \tag{4}$$

where $[\cdot]_+ \triangleq \max(0, \cdot)$, $\mathcal{KG}^-$ contains incorrect triplets constructed by replacing head entity or tail entity in a valid triplet randomly, and γ controls the margin between positive and negative triplets.

TUP

Inspired by the above translation assumption between two entities in KG, we propose TUP to explictly models user preferences and regards them as translational relationships between users and items. Given a set of user-item interactions $\mathcal{Y}$, it automatically induces a preference for a user-items pair, and learns the embeddings of preference $\mathbf{p}$, user $\mathbf{u}$ and item $\mathbf{i}$, satisfying $\mathbf{u} + \mathbf{p} \approx \mathbf{i}$.

Preference Induction

Given a user-item pair (u, i), this component is to induce a preference from a set of latent factors $\mathcal{P}$. These factors are shared by all users, and each $p \in \mathcal{P}$ denotes a different preference, which aims at capturing the commonality among users as global features complementary to user embeddings that focus on a single user locally. Similar with topic models, the number $P = |\mathcal{P}|$ is a hyperparameter, and we cannot nominate the exact meaning of each preference. With the help of KG, the number of preferences can be automatically set and each preference is assigned with explanations

Preference Induction

❑ Hard Strategy – Straight Through Gumbel-Softmax

ST Gumbel SoftMax approximately samples one-hot vectors from a multi-classification distribution. Assuming that the probability of belonging to class p in a P-way classification distribution is defined as log softmax:

$$\phi(p) = \frac{\exp(\log(\pi_p))}{\sum_{j=1}^{P} \exp(\log(\pi_j))} \tag{5}$$

where π_p is the unnormalized output of a score function.

Preference Induction

we sample a one-hot vector $\mathbf{z} = [z_1, \cdots, z_P] \in \mathbb{R}^P$ from the above distribution as follows:

$$z_p = \begin{cases} 1, & p = \arg\max_j (\log(\pi_j) + g_j) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $g = -\log(-\log(u))$ is the Gumbel noise and u is generated by a certain noise distribution (e.g., $u \sim \mathcal{N}(0,1)$). The noise term increases the stochastic of arg max function and makes the process become equivalent to drawing a sample accroding to a continuous probability distribution $\mathbf{y} = [y_1, \cdots, y_p, \cdots, y_P]$:

Gumbel-Softmax Distribution

$$y_p = \frac{\exp((\log(\pi_p) + g_p)/\tau)}{\sum_{j=1}^{P} \exp((\log(\pi_j) + g_j)/\tau)} \tag{7}$$

서울시립대학교 UNIVERSITY OF SEOUL

Preference Induction

In the hard strategy, we define the score function for π_p as the similarity between the user-item pair and preference:

$$\phi(u, i, p) = \text{Similarity}(\mathbf{u} + \mathbf{i}, \mathbf{p}) \tag{8}$$

서울시립대학교
UNIVERSITY OF SEOUL

Preference Induction

❑ Soft Strategy – Attention Mechanism

Instead of selecting the most prominent preference, the soft strategy is to combine multiple preferences via the attention mechanism:

$$\mathbf{p} = \sum_{p' \in \mathcal{P}} \alpha_{p'} \mathbf{p}' \tag{9}$$

where $\alpha_{p'}$ is the attention weight of preference p', and defined as proportional to the similarity score:

$$\alpha_{p'} \propto \phi(u, i, p') \tag{10}$$

Hyperplane-based Translation

$$g(u, i; p) = \| \mathbf{u}^{\perp} + \mathbf{p} - \mathbf{i}^{\perp} \| \tag{11}$$

where $\mathbf{u}^{\perp}$ and $\mathbf{i}^{\perp}$ are projected vectors of the user and the item, and are obtained through the induced preference p that plays a similar role as relations in TransH:

$$\mathbf{u}^{\perp} = \mathbf{u} - \mathbf{w}_p^{\mathrm{T}} \mathbf{u} \mathbf{w}_p \tag{12}$$

$$\mathbf{i}^{\perp} = \mathbf{i} - \mathbf{w}_p^{\mathrm{T}} \mathbf{i} \mathbf{w}_p \tag{13}$$

Hyperplane-based Translation

$$\mathbf{w}_p = \sum_{p' \in \mathcal{P}} \alpha_{p'} \mathbf{w}_{p'} \tag{14}$$

We encourage the translation distances of the interacted items to be smaller than random ones for each user through BPR Loss function:

$$\mathcal{L}_p = \sum_{(u,i) \in \mathcal{Y}} \sum_{(u,i') \in \mathcal{Y}'} - \log \sigma [g(u, i'; p') - g(u, i; p)] \tag{15}$$

where $\mathcal{Y}'$ contains negative interactions by randomly corrupting an interacted item to a non-interacted one for each user.

서울시립대학교
UNIVERSITY OF SEOUL

KTUP

$$g(u, i; p) = \| \mathbf{u}^{\perp} + \hat{\mathbf{p}} - \hat{\mathbf{i}}^{\perp} \| \tag{16}$$

where $\hat{\mathbf{i}}^{\perp}$ is the projected vector for the enhanced item embedding $\hat{\mathbf{i}}$ by the corresponding entity embedding $\mathbf{e}$:

$$\hat{\mathbf{i}}^{\perp} = \hat{\mathbf{i}} - \hat{\mathbf{w}}_p^{\mathrm{T}} \hat{\mathbf{i}} \hat{\mathbf{w}}_p \tag{17}$$

$$\hat{\mathbf{i}} = \mathbf{i} + \mathbf{e}, \ (i, e) \in \mathcal{A} \tag{18}$$

And $\hat{\mathbf{p}}$ and $\hat{\mathbf{w}}_p$ are the translation vector and the projection vector enhanced by those of the corresponding relation embedding according a predefined one-to-one mapping $\mathcal{R} \rightarrow \mathcal{P}$. We obtain these two vectors as follows:

$$\hat{\mathbf{p}} = \mathbf{p} + \mathbf{r} \tag{19}$$

$$\hat{\mathbf{w}}_p = \mathbf{w}_p + \mathbf{w}_r \tag{20}$$

KTUP - Training

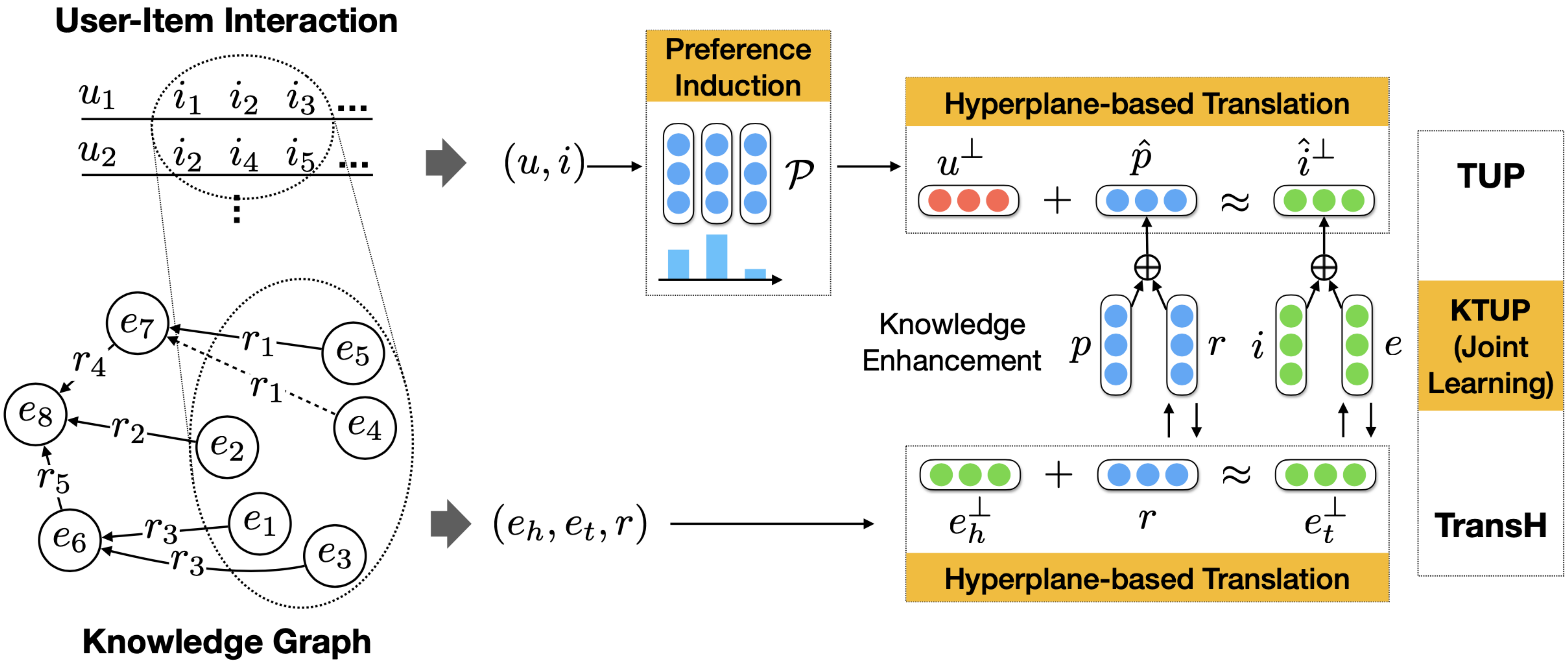
$$\mathcal{L} = \lambda \mathcal{L}_p + (1 - \lambda)\mathcal{L}_k$$

$$\mathcal{L}_p = \sum_{(u,i)\in\mathcal{Y}} \sum_{(u,i')\in\mathcal{Y}'} -\log \sigma[g(u,i';p') - g(u,i;p)]$$

$$\mathcal{L}_k = \sum_{(e_h,e_t,r)\in\mathcal{KG}} \sum_{(e'_h,e'_t,r')\in\mathcal{KG}^-} [f(e_h,e_t,r) + \gamma - f(e'_h,e'_t,r')]_+$$

서울시립대학교
UNIVERSITY OF SEOUL

Framework of KTUP



서울시립대학교
UNIVERSITY OF SEOUL

Datasets

		MovieLens-1m	DBbook2014
User-Item Interactions	# Users	6,040	5,576
	# Items	3,240	2,680
	# Ratings	998,539	65,961
	# Avg. ratings	165	12
	Sparsity	94.9%	99.6%
KG	# Entity	14,708	13,882
	# Relation	20	13
	# Triple	434,189	334,511
Multi-Tasks	# Item-Entity Alignments	2,934	2,534
	Coverage	90.6%	94.6%

Experiments

❑ Performance on Item Recommendation

	MovieLens-1m (@10, %)					DBbook2014 (@10, %)				
	Precision	**Recall**	**F1**	**Hit**	**NDCG**	**Precision**	**Recall**	**F1**	**Hit**	**NDCG**
FM	29.28	11.92	13.81	81.06	59.48	3.44	21.55	5.75	30.15	20.10
BPRMF	30.81	12.95	14.84	83.18	61.02	3.56	22.46	5.96	31.26	21.01
CFKG	29.45	12.49	14.23	82.24	58.97	3.17	19.69	5.30	28.09	19.87
CKE	38.67	16.65	18.94	88.36	67.05	3.92	23.41	6.51	33.18	**27.78**
CoFM (share)	32.08	13.02	15.12	83.30	58.69	3.41	20.78	5.67	29.84	20.92
CoFM (reg)	31.74	12.74	14.87	82.67	58.66	3.32	20.54	5.54	28.96	20.53
TUP (hard)	37.29	17.07	18.98	**89.60**	67.40	3.40	21.11	5.67	29.56	20.19
TUP (soft)	37.00	16.79	18.76	89.47	67.02	3.62	22.81	6.06	31.42	21.54
KTUP (hard)	40.87	17.24	19.79	88.97	69.65	4.04	24.48	6.71	34.49	27.38
KTUP (soft)	**41.03**	**17.25**	**19.82**	89.03	**69.92**	**4.05**	**24.51**	**6.73**	**34.61**	27.62

Experiments

❑ Performance on MovieLens-1m by Relation Category

Task	Prediction Head (Hits@10, %)				Prediction Tail (Hits@10, %)			
Relation Category	**1-to-1**	**1-to-N**	**N-to-1**	**N-to-N**	**1-to-1**	**1-to-N**	**N-to-1**	**N-to-N**
TransE	59.62	56.76	64.55	24.56	**65.38**	62.16	78.52	46.25
TransH	61.54	48.65	65.73	25.51	57.69	78.38	75.62	46.73
TransR	17.31	29.73	32.88	18.50	17.31	43.24	53.12	38.88
CFKG	59.62	51.35	63.31	20.30	57.69	70.27	78.56	41.22
CKE	19.23	21.62	24.16	14.81	7.69	24.32	37.83	34.82
CoFM (share)	65.38	59.46	66.13	24.42	61.54	72.97	**81.05**	45.99
CoFM (reg)	69.23	**70.27**	66.09	24.30	48.08	**86.49**	80.72	45.79
KTUP (hard)	67.31	59.46	66.42	25.67	57.69	81.08	79.22	47.24
KTUP (soft)	**75.00**	56.76	**67.16**	**26.09**	63.46	81.08	78.34	**47.65**

Experiments