当知识图谱遇见个性化推荐

张富峥
微软亚洲研究院
社会计算组 研究员

https://www.microsoft.com/en-us/research/people/fuzzhang/
fuzzhang@microsoft.com

Microsoft

个性化推荐的广泛应用

电影（豆瓣）

图书（亚马逊）

音乐（网易云音乐）

餐馆（大众点评）

旅游（飞猪）

新闻（今日头条）

个性化推荐的任务

Top-K 推荐

2		3	
		4	
	5		2
3	1	4	

你还可能喜欢...

战狼
拯救大兵瑞恩
血战钢锯岭
阿甘正传

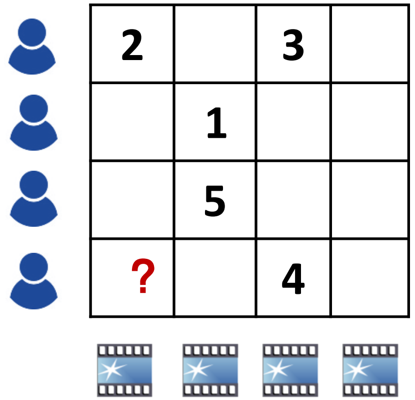
协同过滤

矩阵分解

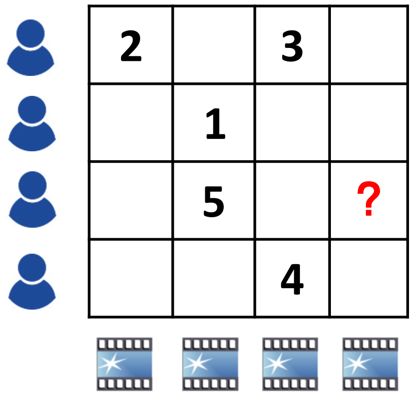
$$L = \| \boldsymbol{R} - \boldsymbol{P}^{\mathrm{T}}\boldsymbol{Q} \|_2^2 + \| \boldsymbol{P} \|_2^2 + \| \boldsymbol{Q} \|_2^2$$

- 仅使用评分信息 / 交互信息
- 无法解决稀疏性和冷启动问题

评分的稀疏性



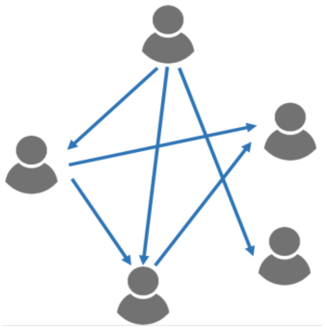
冷启动问题（新用户/物品）



协同过滤+辅助信息

- 在推荐系统中融合以下辅助信息：

社交网络



用户/物品属性

Alice
Female
California
…

iPhone X
2017
5.8 inch
$999
…

图片/文字 等内容信息

A great blend of handheld comfort and a big, gorgeous OLED screen. Rear telephoto camera outshoots the 8 Plus in low light, and the front camera snaps impressive portrait selfies. Face ID generally works fine

上下文

购买

时间: 20点10分
定位：北京
购物车内还有：...

知识图谱

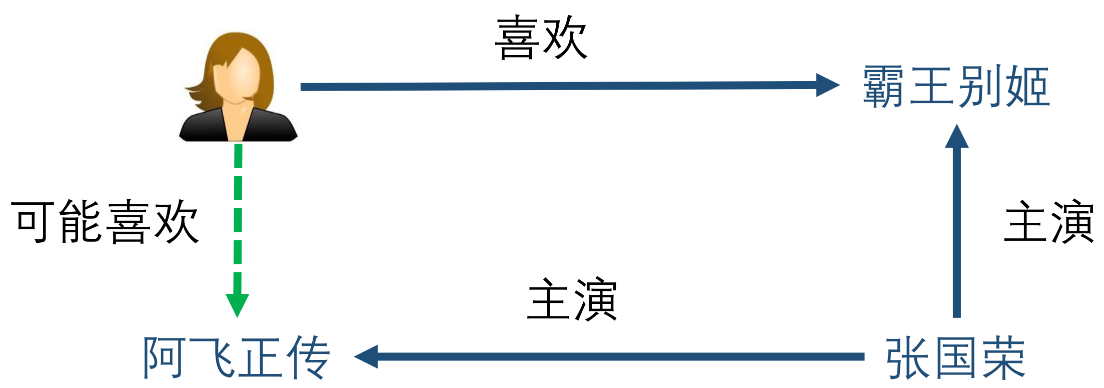
知识图谱（knowledge graph）是一种语义网络，其结点代表实体（entity）或者概念（concept），边代表实体/概念之间的各种语义关系（relation）

- 由三元组（h, r, t）组成，其中h和t代表一条关系的头结点和尾节点，r代表关系
- 例如： 陈凯歌 $\xrightarrow{\text{导演}}$ 霸王别姬
- 知识图谱为推荐系统提供了物品之间丰富的语义关系

知识图谱与推荐系统

引入知识图谱可以让推荐结果更具有：

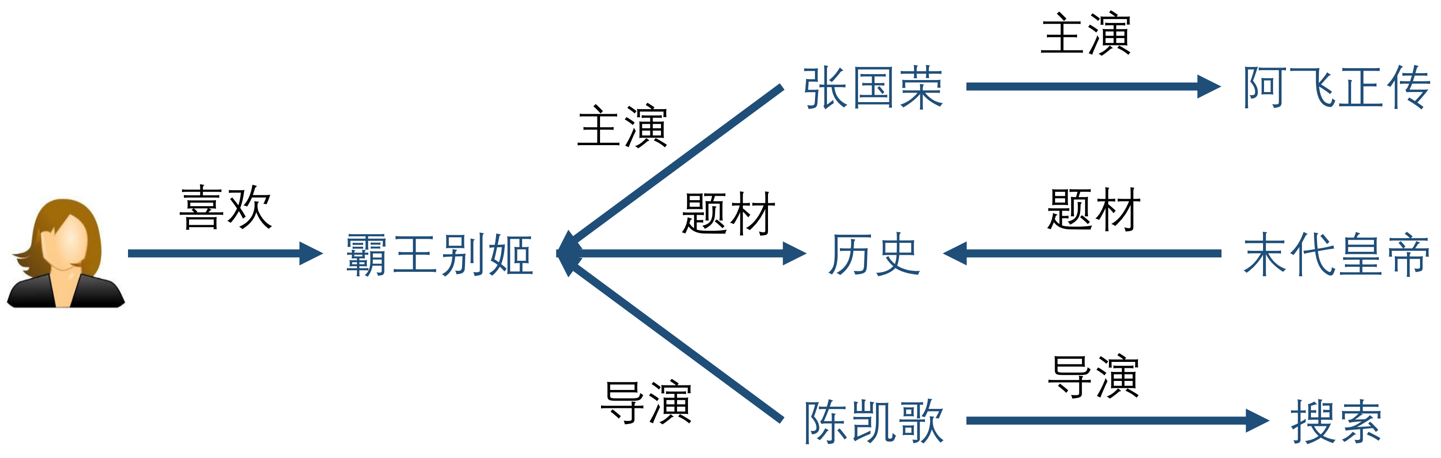
- 精确性（precision）
  - 知识图谱为物品引入了更多的语义关系
  - 知识图谱可以深层次地发现用户兴趣



知识图谱与推荐系统

引入知识图谱可以让推荐结果更具有：

- 多样性（diversity）
  - 知识图谱提供了不同的关系连接种类
  - 有利于推荐结果的发散，避免推荐结果越来越局限于单一类型



知识图谱与推荐系统

引入知识图谱可以让推荐结果更具有：

- 可解释性（explanation ability）
  - 知识图谱可以连接用户的兴趣历史和推荐结果
  - 提高用户对推荐结果的满意度和接受度，增强用户对推荐系统的信任

喜欢

霸王别姬

你可能也喜欢：

阿飞正传，因为它们有相同的主演；
末代皇帝，因为它们有相同的题材；
搜索，因为它们有相同的导演；
……

相关工作

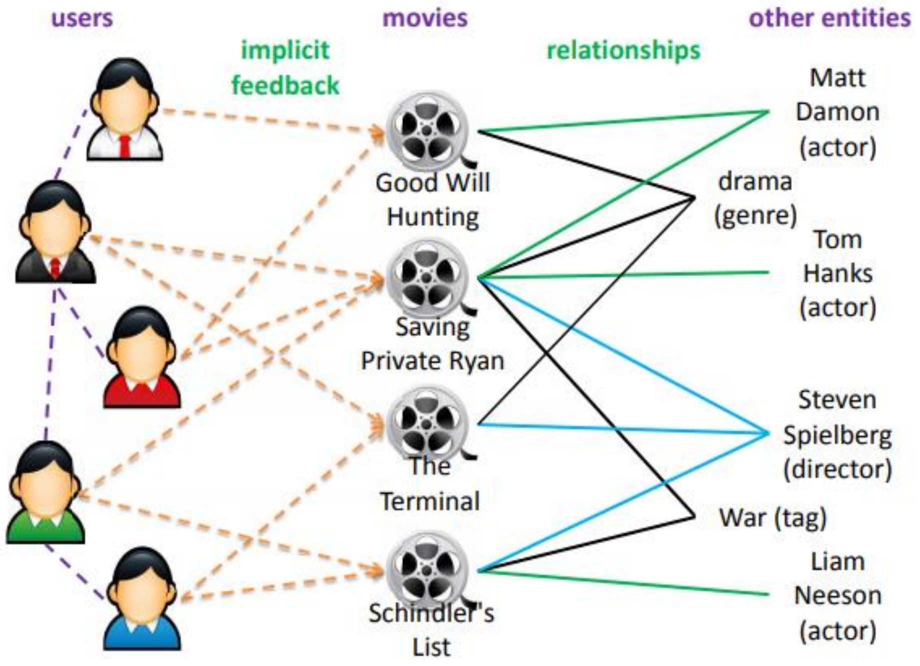
LibFM (TIST 12)

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{j=1}^{p} w_j x_j + \sum_{j=1}^{p} \sum_{j'=j+1}^{p} x_j x_{j'} \sum_{f=1}^{k} v_{j,f}\, v_{j',f}$$

- 优点：
  - 通用的基于特征的推荐方法
  - 模拟特征间的交互行为
- 缺点：
  - 并非专门针对知识图谱设计
  - 无法引入关系（relation）特征

PER (WSDM 14)
MetaGraph （KDD 17)

- 优点：
  - 构造物品间的metapath/metagraph
  - 对知识图谱的使用更简洁直观
- 缺点：
  - 需要手动设计metapath/metagraph，实践中难以达到最优
  - 在实体不属于同一个领域的场景中（如新闻推荐）无法应用



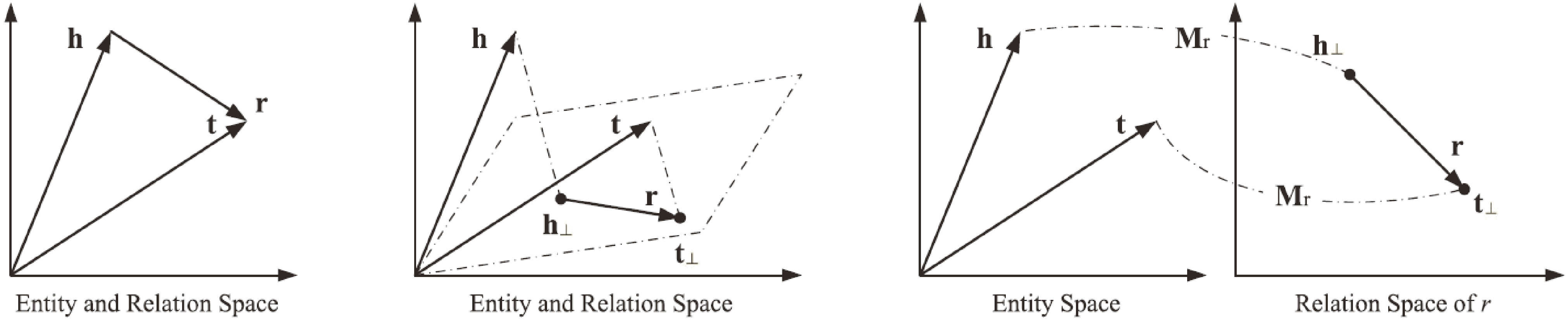
meta path 1: movie -> actor -> movie
meta path 2: movie -> director -> movie

知识图谱特征学习

知识图谱特征学习（Knowledge Graph Embedding）为知识图谱中的每个实体和关系学习得到一个低维向量，同时保持图中原有的结构或语义信息

基于距离的翻译模型

- 使用基于**距离**的评分函数评估三元组的概率
- 将尾节点视为头结点和关系翻译得到的结果
- TransE, TransH, TransR等



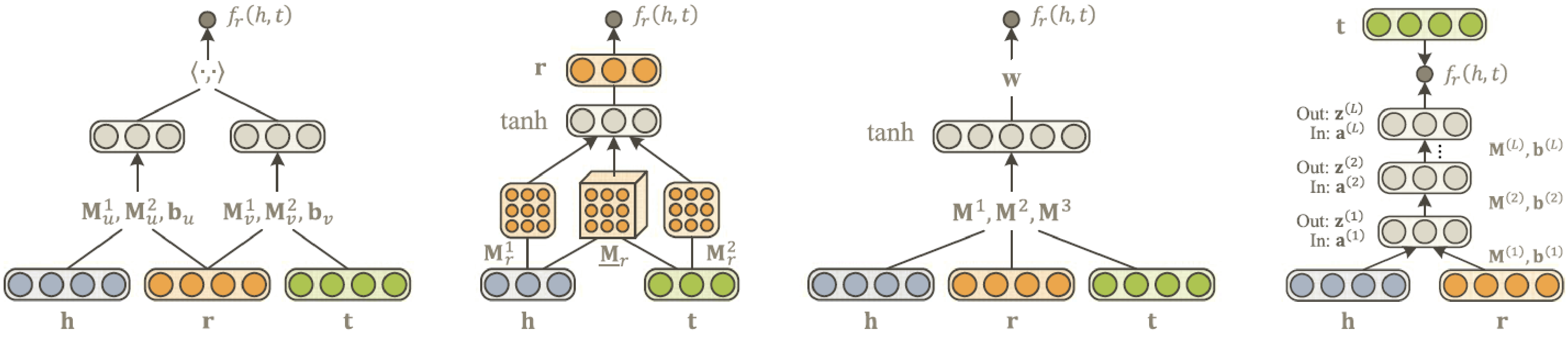
(a) TransE. (b) TransH. (c) TransR.

知识图谱特征学习

知识图谱特征学习（Knowledge Graph Embedding）为知识图谱中的每个实体和关系学习得到一个低维向量，同时保持图中原有的结构或语义信息

基于语义的匹配模型

- 使用基于**相似度**的评分函数评估三元组的概率
- 将实体和关系映射到隐含语义空间进行相似度度量
- SME, NTN, MLP, NAM等



(a) SME. (b) NTN. (c) MLP. (d) NAM.

结合知识图谱特征学习的推荐系统

知识图谱特征学习可以：

- 降低知识图谱的高维性和异构性
- 增强知识图谱应用的灵活性
- 减轻特征工程的工作量
- 减少由于引入知识图谱带来的额外计算负担

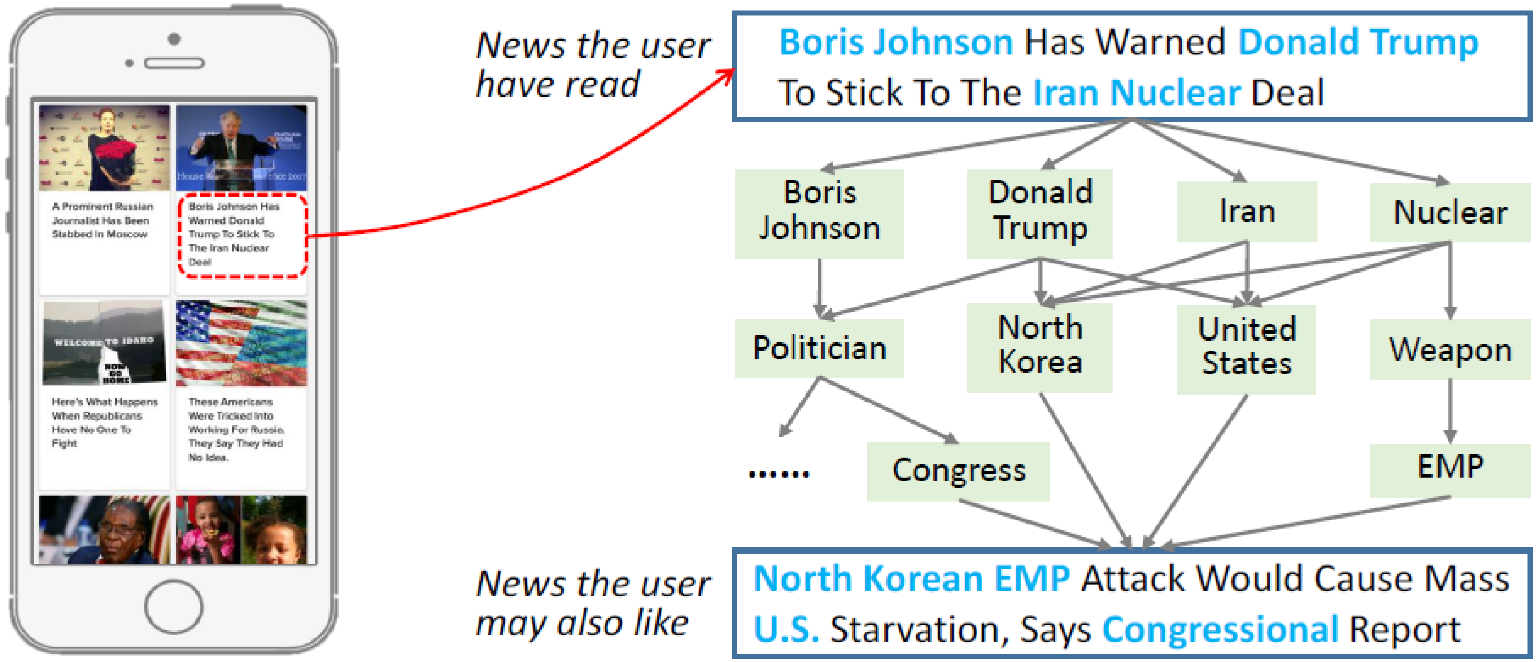
知识图谱特征学习在推荐系统中的应用步骤：

- 知识图谱 $\xrightarrow{\text{KGE}}$ 实体向量 关系向量 $\xrightarrow{\text{引入}}$ 推荐系统 $\xrightarrow{\text{学习}}$ 用户向量 物品向量 **（依次训练）**
- 知识图谱 $\xrightarrow{\text{KGE}}$ / 推荐系统 $\xrightarrow{\text{学习}}$ → 实体向量、关系向量 用户向量、物品向量 **（联合训练）**
- 知识图谱 $\xrightarrow{\text{KGE}}$ 实体向量、关系向量
  推荐系统 $\xrightarrow{\text{学习}}$ 用户向量、物品向量 **（交替训练）**

依次训练

Deep Knowledge-aware Network (DKN)

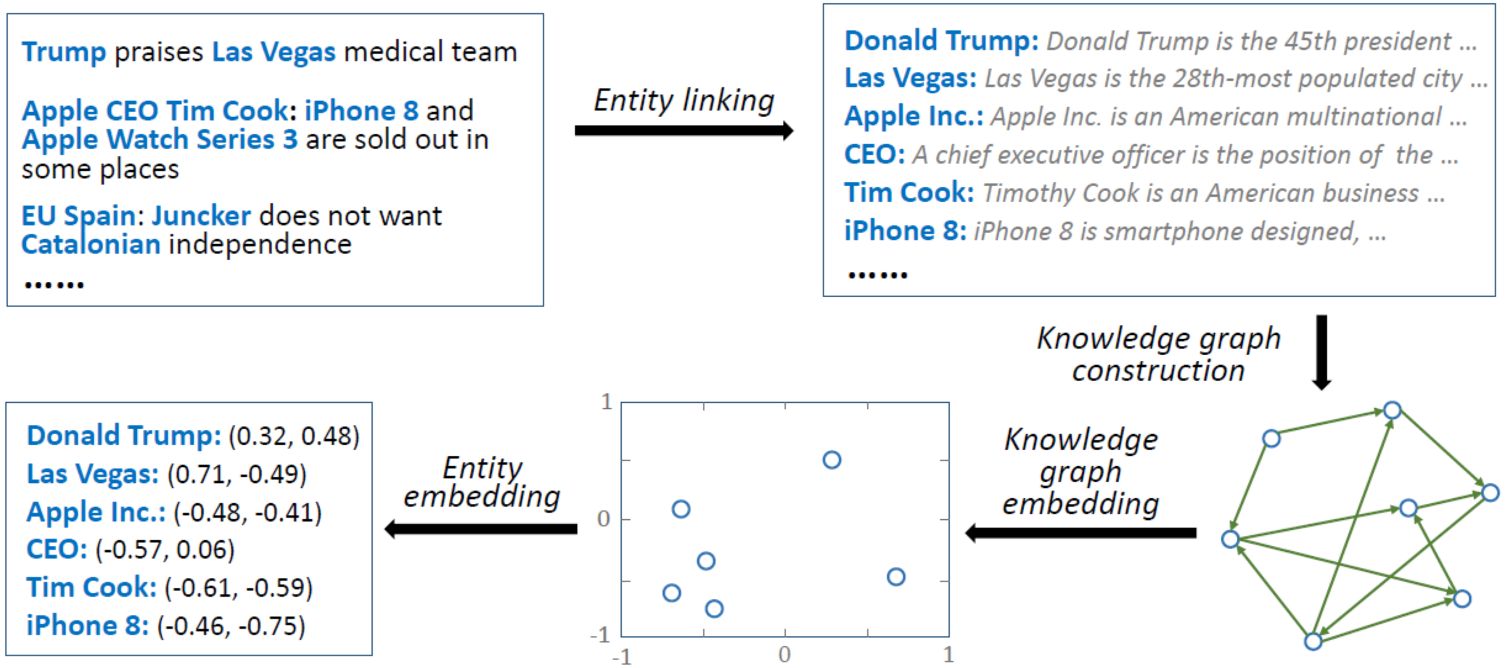
- 新闻标题和正文中存在大量的实体
- 实体间的语义关系可以有效地扩展用户兴趣
- 实体间的语义关系难以被传统方法（话题模型、词向量方法）发掘



依次训练

Deep Knowledge-aware Network (DKN)

- 实体连接（entity linking）
- 知识图谱构建（knowledge graph construction）
- 知识图谱特征学习（knowledge graph embedding）
- 得到实体特征（entity embedding）

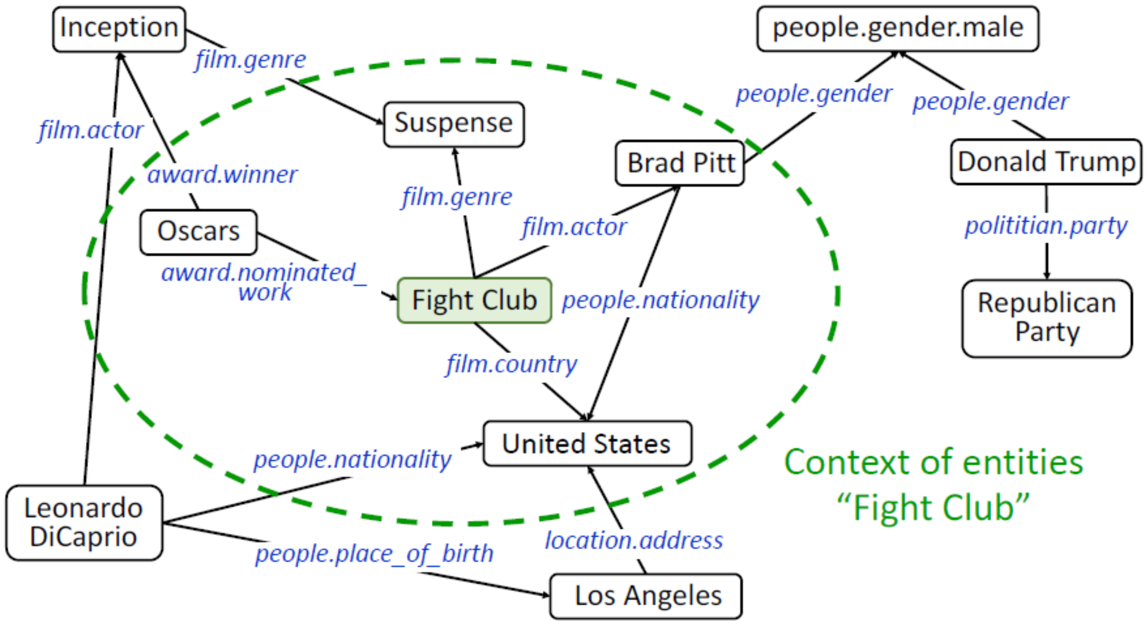


依次训练

Deep Knowledge-aware Network (DKN)

第一步：提取知识图谱特征

- ❑ 额外使用一个实体的上下文实体特征（contextual entity embeddings）对该实体进行更准确地刻画
- ❑ 上下文实体为该实体的一度邻居节点



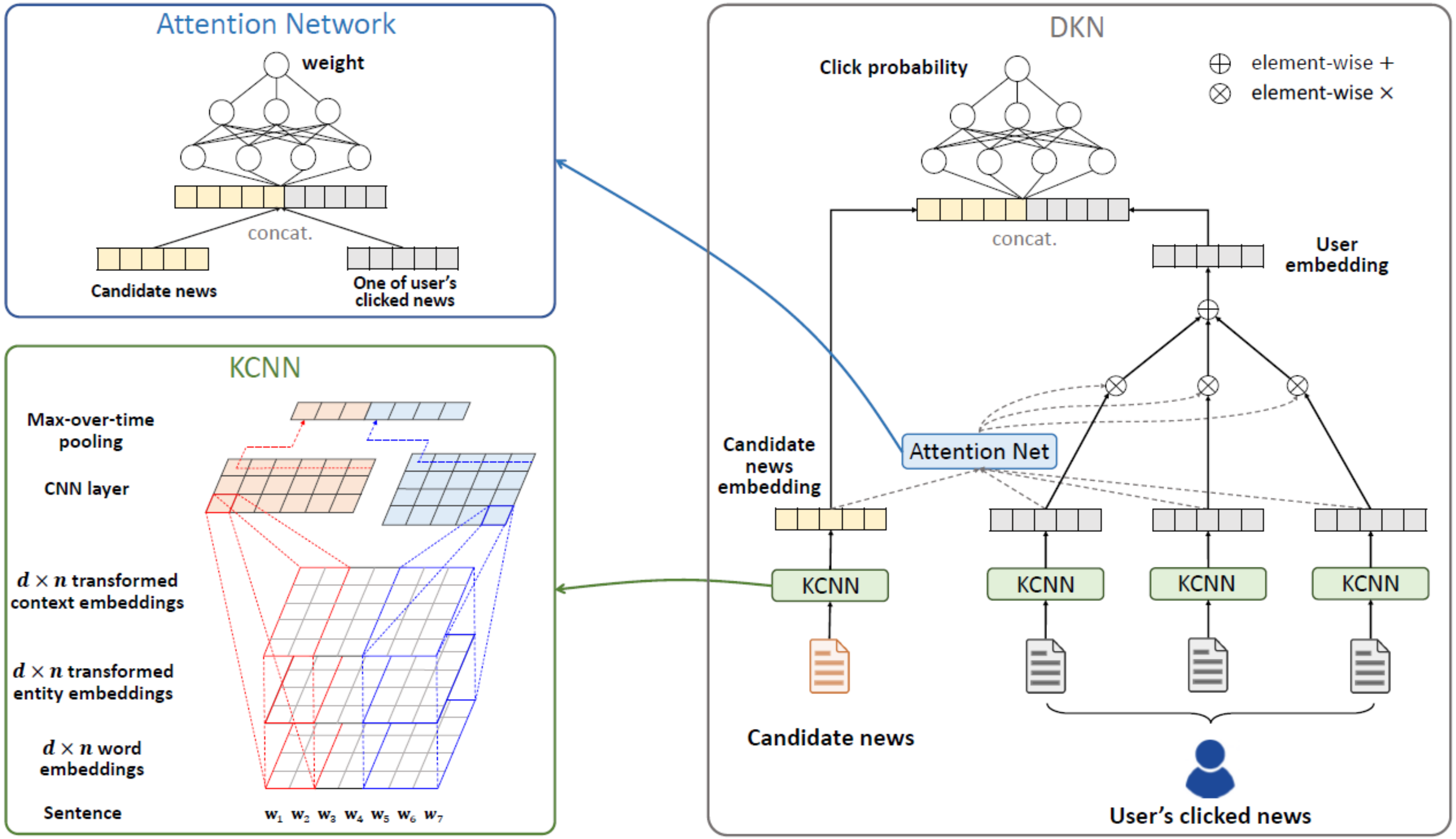
$$context(e) = \{e_i \mid (e, r, e_i) \in \mathcal{G} \text{ or } (e_i, r, e) \in \mathcal{G}\}$$

$$\bar{\mathbf{e}} = \frac{1}{|context(e)|} \sum_{e_i \in context(e)} \mathbf{e}_i,$$

依次训练

Deep Knowledge-aware Network (DKN)

第二步：构建推荐模型



依次训练

Deep Knowledge-aware Network (DKN)

第二步：构建推荐模型

- 基于CNN的文本特征提取
- 词向量、实体向量、实体上下文向量的多通道融合

news title $t = [w_1, w_2, \ldots, w_n]$

word embeddings $\mathbf{w}_{1:n} = [\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_n]$

transformed entity embeddings

$$g(\mathbf{e}_{1:n}) = [g(\mathbf{e}_1)\ g(\mathbf{e}_2)\ \ldots\ g(\mathbf{e}_n)]$$

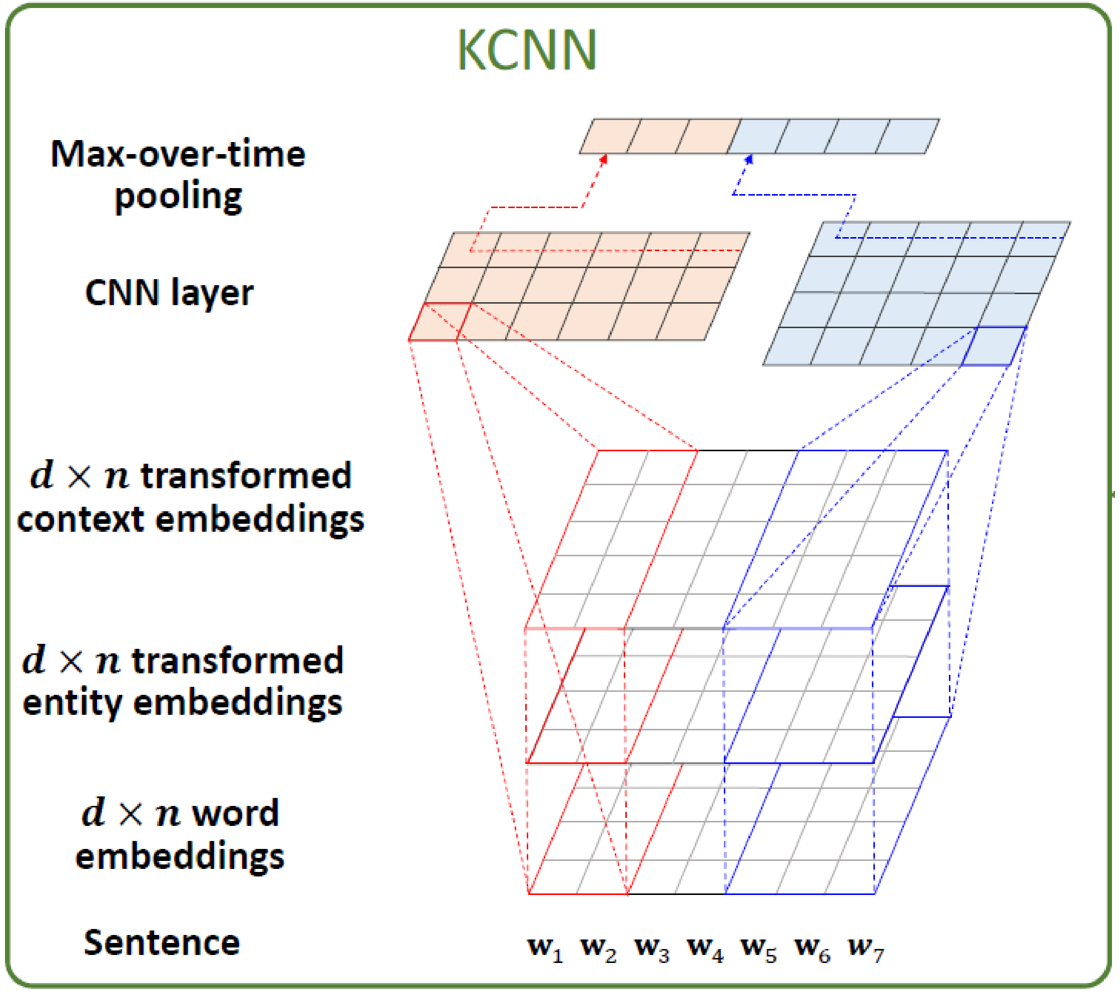
transformed context embeddings

$$g(\overline{\mathbf{e}}_{1:n}) = [g(\overline{\mathbf{e}}_1)\ g(\overline{\mathbf{e}}_2)\ \ldots\ g(\overline{\mathbf{e}}_n)]$$

$$\mathbf{W} = [[\mathbf{w}_1\ g(\mathbf{e}_1)\ g(\overline{\mathbf{e}}_1)]\ [\mathbf{w}_2\ g(\mathbf{e}_2)\ \overline{g}(\mathbf{e}_2)]\ \ldots\ [\mathbf{e}_n\ g(\mathbf{e}_n)\ g(\overline{\mathbf{e}}_n)]] \in \mathbb{R}^{d\times n\times 3}$$

$$c_i^h = f(\mathbf{h} * \mathbf{W}_{i:i+l-1} + b) \qquad \tilde{c}^h = \max\{c_1^h, c_2^h, \ldots, c_{n-l+1}^h\}$$

$$\mathbf{e}(t) = [\tilde{c}^{h_1}\ \tilde{c}^{h_2}\ \ldots\ \tilde{c}^{h_m}]$$



依次训练

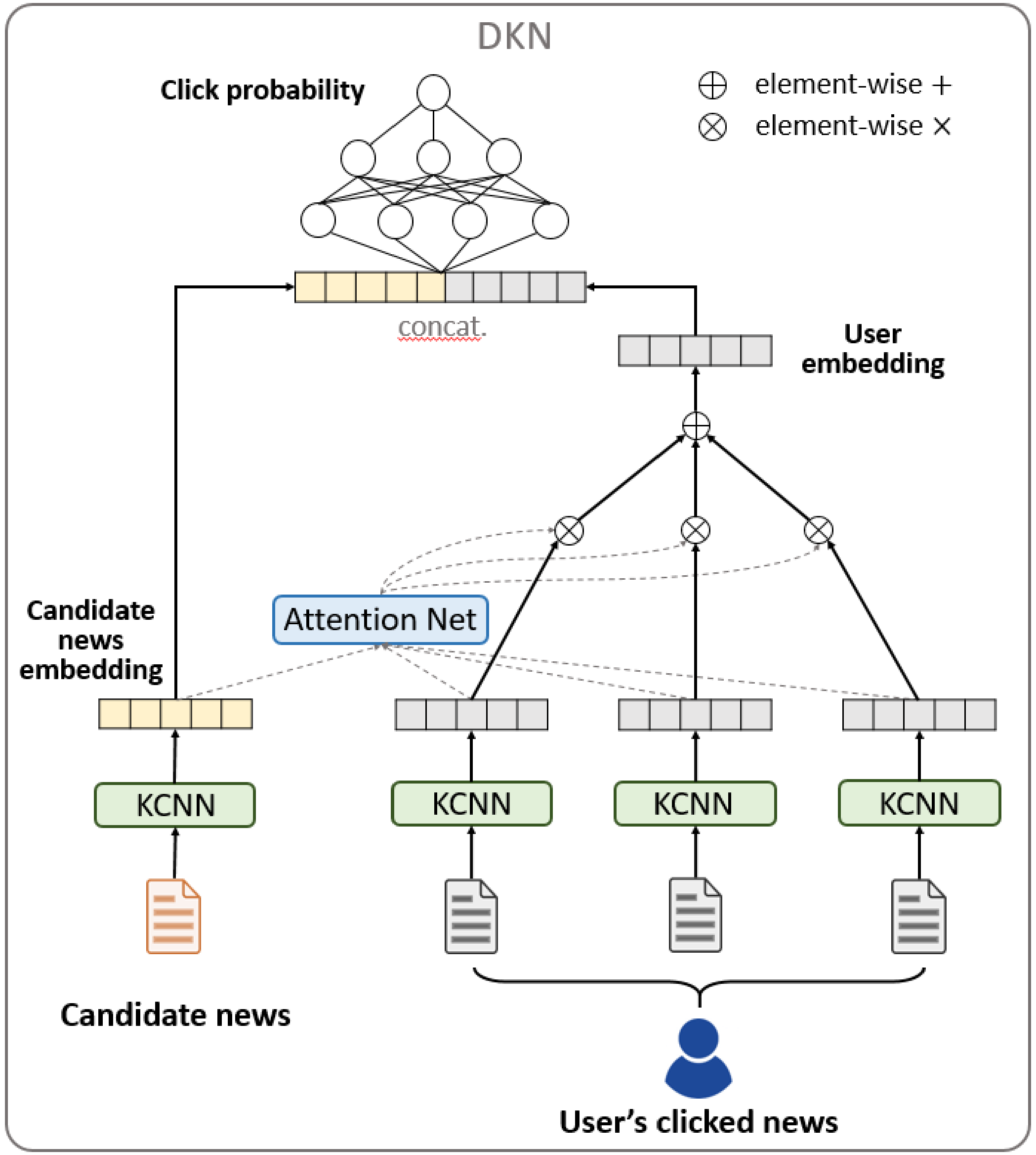
Deep Knowledge-aware Network (DKN)

第二步：构建推荐模型

- 基于注意力机制的用户历史兴趣融合
- 判断用户对当前新闻的兴趣时，用户历史中不同的行为记录的重要性有所不同
- 注意力网络（attention net）可以给用户历史的记录分配不同的权值

$$s_{t_k^i, t_j} = \mathrm{softmax}\left(\mathcal{H}\left(\mathbf{e}(t_k^i), \mathbf{e}(t_j)\right)\right) = \frac{\exp\left(\mathcal{H}\left(\mathbf{e}(t_k^i), \mathbf{e}(t_j)\right)\right)}{\sum_{k=1}^{N_i} \exp\left(\mathcal{H}\left(\mathbf{e}(t_k^i), \mathbf{e}(t_j)\right)\right)}$$

$$\mathbf{e}(i) = \sum_{k=1}^{N_i} s_{t_k^i, t_j} \mathbf{e}(t_k^i)$$



依次训练

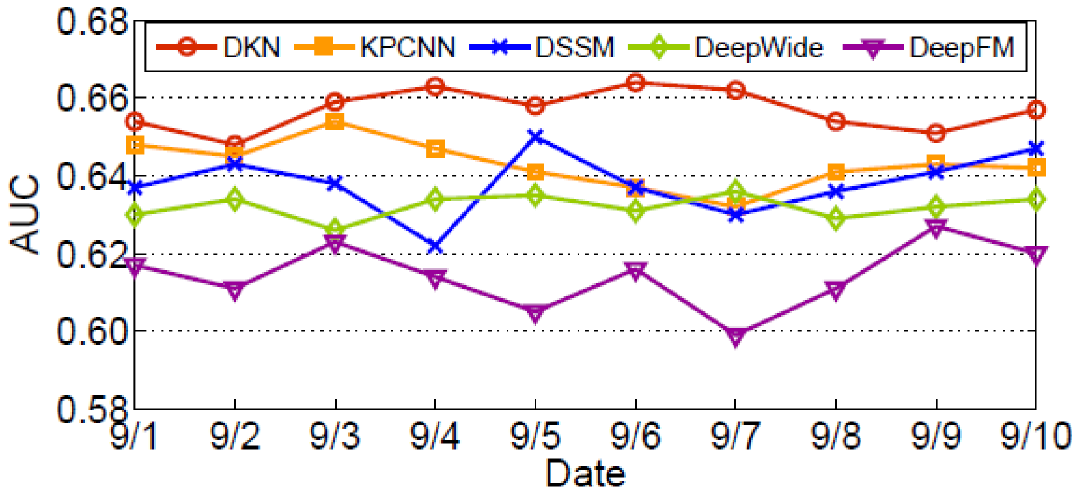
Deep Knowledge-aware Network (DKN)

性能对比

Models*	F1	AUC	p-value**
DKN	**68.9 ± 1.5**	**65.9 ± 1.2**	–
LibFM	61.8 ± 2.1 (-10.3%)	59.7 ± 1.8 (-9.4%)	$< 10^{-3}$
LibFM(-)	61.1 ± 1.9 (-11.3%)	58.9 ± 1.7 (-10.6%)	$< 10^{-3}$
KPCNN	67.0 ± 1.6 (-2.8%)	64.2 ± 1.4 (-2.6%)	0.098
KPCNN(-)	65.8 ± 1.4 (-4.5%)	63.1 ± 1.5 (-4.2%)	0.036
DSSM	66.7 ± 1.8 (-3.2%)	63.6 ± 2.0 (-3.5%)	0.063
DSSM(-)	66.1 ± 1.6 (-4.1%)	63.2 ± 1.8 (-4.1%)	0.045
DeepWide	66.0 ±1.2 (-4.2%)	63.3 ± 1.5 (-3.9%)	0.039
DeepWide(-)	63.7 ± 0.9 (-7.5%)	61.5 ± 1.1 (-6.7%)	0.004
DeepFM	63.8 ± 1.5 (-7.4%)	61.2 ± 2.3 (-7.1%)	0.014
DeepFM(-)	64.0 ± 1.9 (-7.1%)	61.1 ± 1.8 (-7.3%)	0.007
YouTubeNet	65.5 ± 1.2 (-4.9%)	63.0 ± 1.4 (-4.4%)	0.025
YouTubeNet(-)	65.1 ± 0.7 (-5.5%)	62.1 ± 1.3 (-5.8%)	0.011
DMF	57.2 ± 1.2 (-17.0%)	55.3 ± 1.0 (-16.1%)	$< 10^{-3}$

* "(-)" denotes "without input of entity embeddings".

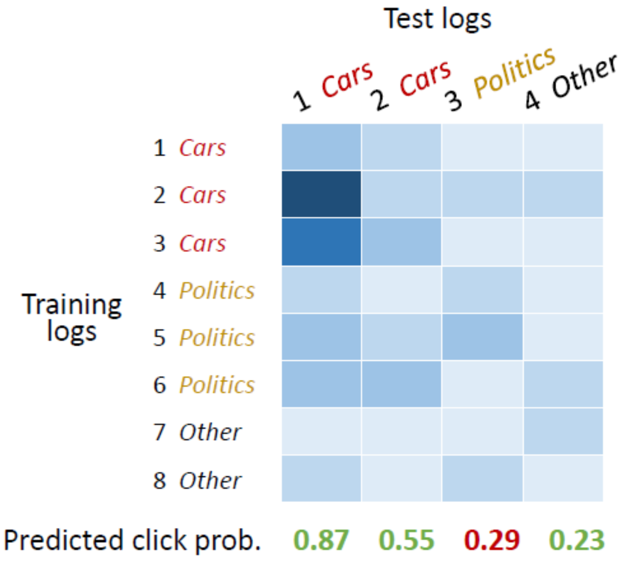
** p-value is the probability of no significant difference with DKN on AUC by t-test.



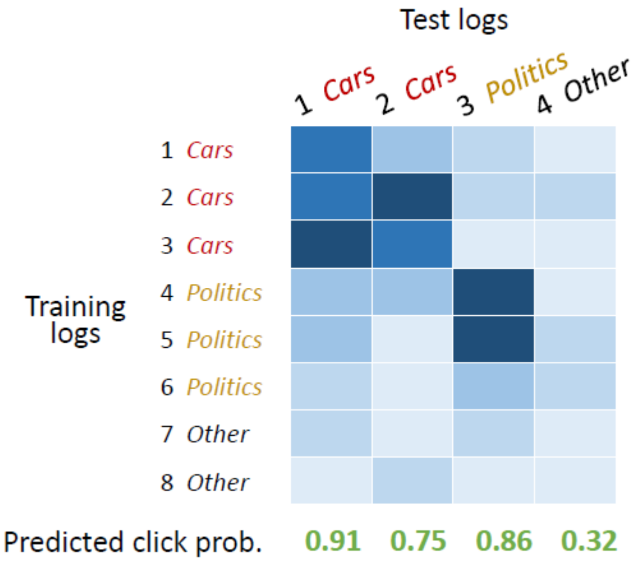
依次训练

Deep Knowledge-aware Network (DKN)

	No.	Date	News title	Entities	Label	Category
training	1	12/25/2016	Elon Musk teases huge upgrades for Tesla's supercharger network	Elon Musk; Tesla Inc.	1	Cars
	2	03/25/2017	Elon Musk offers Tesla Model 3 sneak peek	Elon Musk; Tesla Model 3	1	Cars
	3	12/14/2016	Google fumbles while Tesla sprints toward a driverless future	Google Inc.; Tesla Inc.	1	Cars
	4	12/15/2016	Trump pledges aid to Silicon Valley during tech meeting	Donald Trump; Silicon Valley	1	Politics
	5	03/26/2017	Donald Trump is a big reason why the GOP kept the Montana House seat	Donald Trump; GOP; Montana	1	Politics
	6	05/03/2017	North Korea threat: Kim could use nuclear weapons as "blackmail"	North Korea; Kim Jong-un	1	Politics
	7	12/22/2016	Microsoft sells out of unlocked Lumia 950 and Lumia 950 XL in the US	Microsoft; Lumia; United States	1	Other
	8	12/08/2017	6.5 magnitude earthquake recorded off the coast of California	earthquake; California	1	Other
			......			
test	1	07/08/2017	Tesla makes its first Model 3	Tesla Inc; Tesla Model 3	1	Cars
	2	08/13/2017	General Motors is ramping up its self-driving car: Ford should be nervous	General Motors; Ford Inc.	1	Cars
	3	06/21/2017	Jeh Johnson testifies on Russian interference in 2016 election	Jeh Johnson; Russian	1	Politics
	4	07/16/2017	"Game of Thrones" season 7 premiere: how you can watch	Game of Thrones	0	Other



(a) without knowledge graph



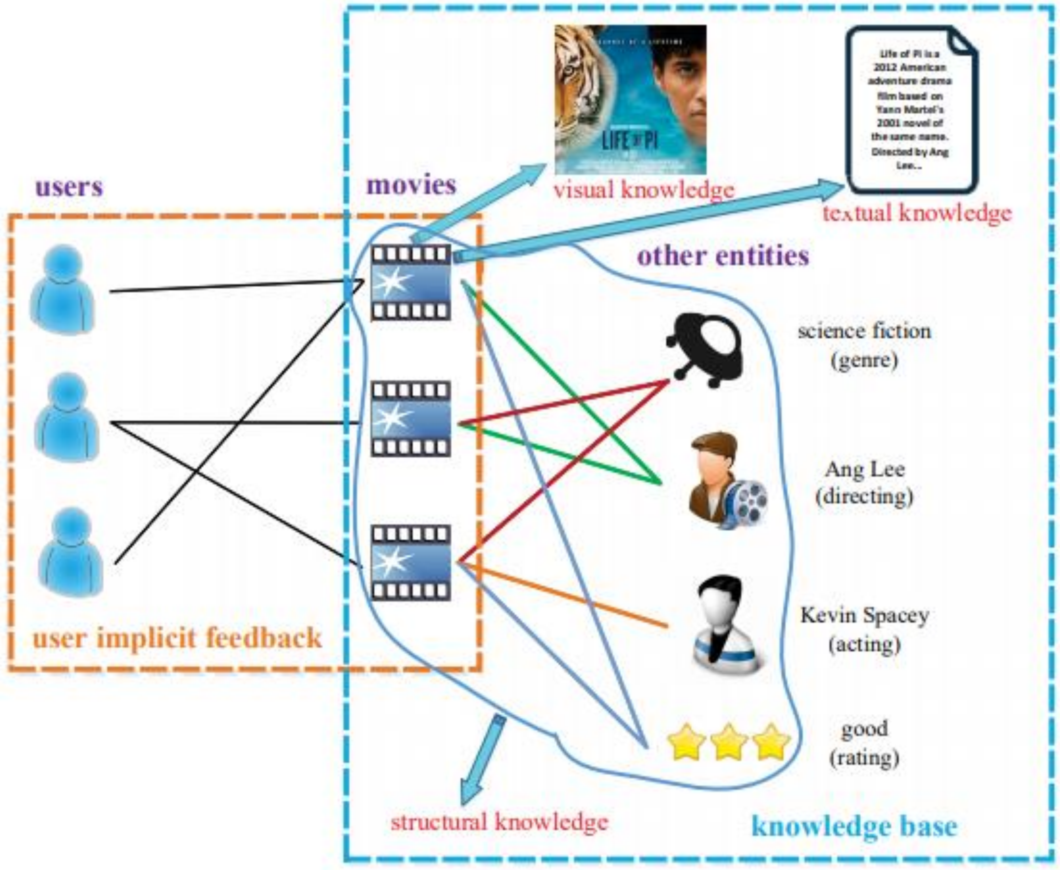
(b) with knowledge graph

联合训练

Collaborative Knowledge base Embedding (CKE)

推荐系统中存在的知识图谱相关信息：

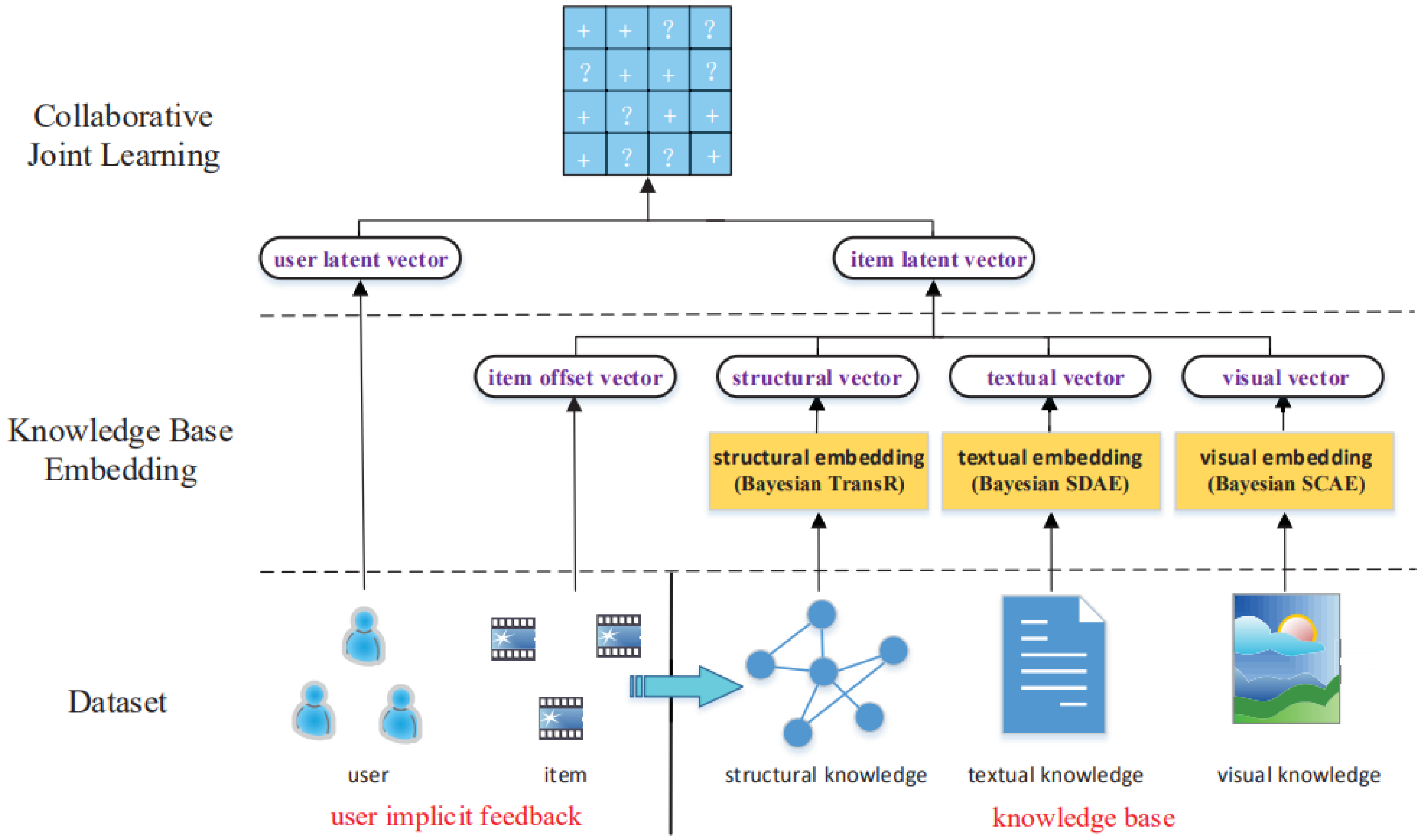
- 结构化知识（structural knowledge）
  - 导演、类别等
- 图片知识（visual knowledge）
  - 海报、剧照等
- 文本知识（textual knowledge）
  - 电影描述、影评等



联合训练

Collaborative Knowledge base Embedding (CKE)

基于协同过滤和知识图谱特征学习的推荐系统

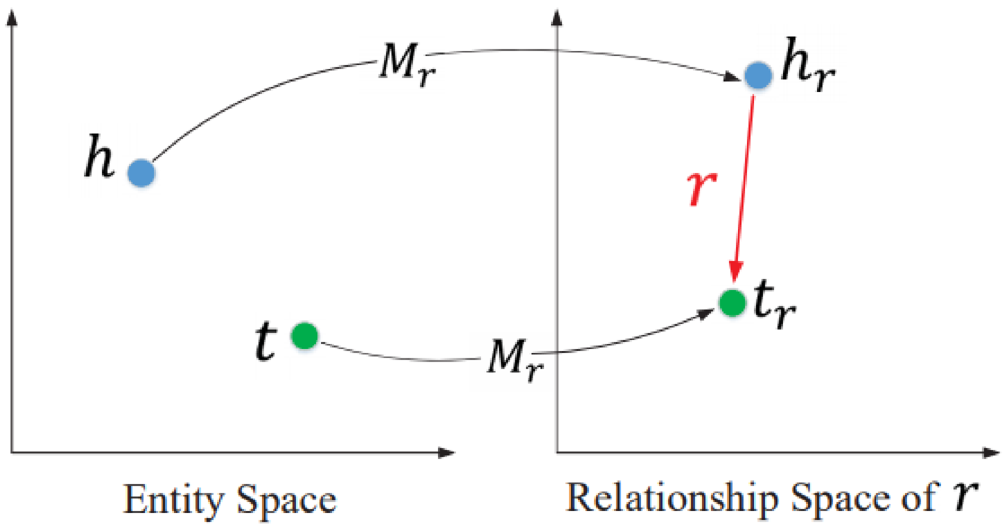


联合训练

Collaborative Knowledge base Embedding (CKE)

结构化知识学习：TransR

1. For each entity v, draw $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \lambda_v^{-1}\mathbf{I})$.
2. For each relation r, draw $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \lambda_r^{-1}\mathbf{I})$ and $\mathbf{M}_r \sim \mathcal{N}(\mathbf{0}, \lambda_M^{-1}\mathbf{I})$, respectively.
3. For each quadruple $(v_h, r, v_t, v_{t'}) \in \mathcal{S}$, draw from the probability $\sigma(f_r(v_h, v_t) - f_r(v_h, v_{t'}))$, where $\mathcal{S}$ is the set of quadruples satisfying that (v_h, r, v_t) is a correct triple and $(v_h, r, v_{t'})$ is an incorrect triple. $\sigma(x) := \frac{1}{1+e^{-x}}$ is the logistic sigmoid function.



$$\mathbf{v}_h^r = \mathbf{v}_h \mathbf{M}_r, \qquad \mathbf{v}_t^r = \mathbf{v}_t \mathbf{M}_r$$

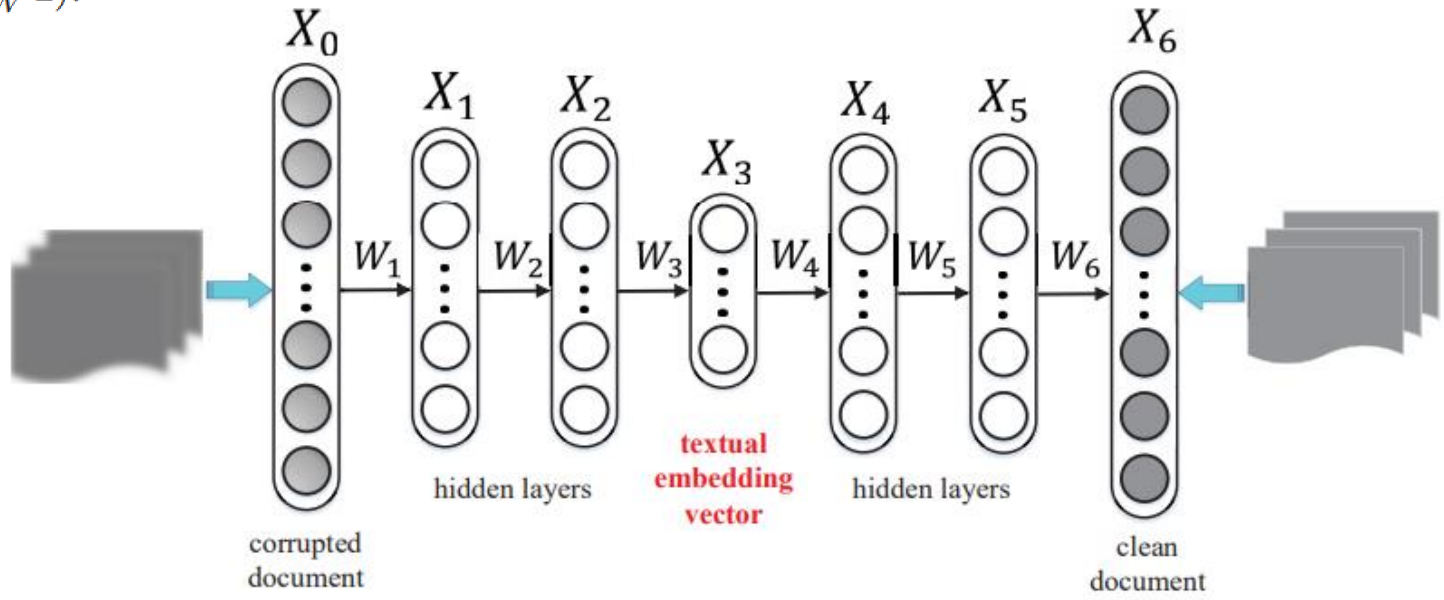
$$f_r(v_h, v_t) = ||\mathbf{v}_h^r + \mathbf{r} - \mathbf{v}_t^r||_2^2$$

联合训练

Collaborative Knowledge base Embedding (CKE)

文本知识学习：去噪自编码器

1. For weight parameter $\mathbf{W}_l$, draw $\mathbf{W}_l \sim \mathcal{N}(\mathbf{0}, \lambda_W^{-1}\mathbf{I})$.
2. For bias parameter, draw $\mathbf{b}_l \sim \mathcal{N}(\mathbf{0}, \lambda_b^{-1}\mathbf{I})$.
3. For the output of the layer, draw $\mathbf{X}_l \sim \mathcal{N}(\sigma(\mathbf{X}_{l-1}\mathbf{W}_l + \mathbf{b}_l), \lambda_X^{-1}\mathbf{I})$

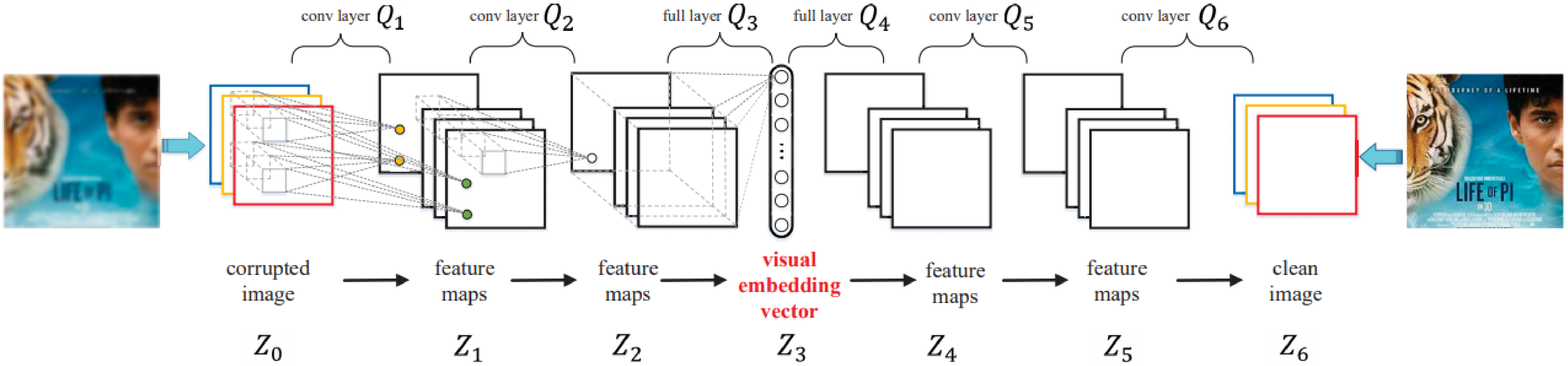


联合训练

Collaborative Knowledge base Embedding (CKE)

图片知识学习：卷积-反卷积自编码器

1. For weight parameter, draw $\mathbf{Q}_l \sim \mathcal{N}(\mathbf{0}, \lambda_Q^{-1}\mathbf{I})$.
2. For bias parameter, draw $\mathbf{c}_l \sim \mathcal{N}(\mathbf{0}, \lambda_c^{-1}\mathbf{I})$.
3. For the output of the layer,
   (a) If layer l is a fully connected layer:
   draw $\mathbf{Z}_l \sim \mathcal{N}(\sigma(\mathbf{Z}_{l-1}\mathbf{Q}_l + \mathbf{c}_l), \lambda_Z^{-1}\mathbf{I})$,
   (b) Else: draw $\mathbf{Z}_l \sim \mathcal{N}(\sigma(\mathbf{Z}_{l-1} * \mathbf{Q}_l + \mathbf{c}_l), \lambda_Z^{-1}\mathbf{I})$.



联合训练

Collaborative Knowledge base Embedding (CKE)

推荐系统和知识图谱特征的联合学习：

4. For each item j, draw a latent item offset vector $\boldsymbol{\eta}_j \sim \mathcal{N}(\mathbf{0}, \lambda_I^{-1}\mathbf{I})$, and then set the item latent vector as: $\mathbf{e}_j = \boldsymbol{\eta}_j + \mathbf{v}_j + \mathbf{X}_{\frac{L_t}{2},j*} + \mathbf{Z}_{\frac{L_v}{2},j*}$.

5. For each user i, draw a user latent vector as $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda_U^{-1}\mathbf{I})$.

6. For each triple $(i, j, j') \in \mathcal{D}$, draw from the probability $\sigma(\mathbf{u}_i^T\mathbf{e}_j - \mathbf{u}_i^T\mathbf{e}_{j'})$.

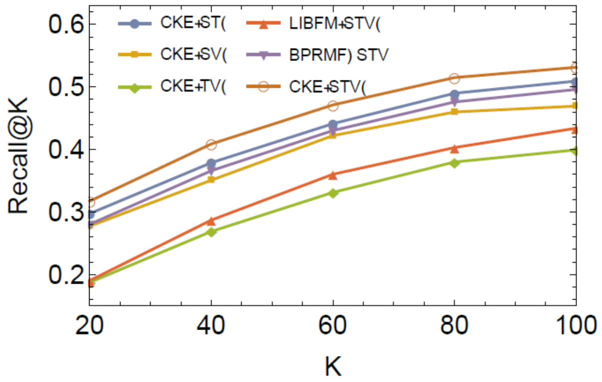
$$
\begin{aligned}
\mathcal{L} = & \sum_{(i,j,j')\in\mathcal{D}} \ln\sigma(\mathbf{u}_i^T\mathbf{e}_j - \mathbf{u}_i^T\mathbf{e}_{j'}) - \frac{\lambda_X}{2}\sum_l ||\sigma(\mathbf{X}_{l-1}\mathbf{W}_l + \mathbf{b}_l) - \mathbf{X}_l||_2^2 \\
& + \sum_{(v_h,r,v_t,v_{t'})\in\mathcal{S}} \ln\sigma(||\mathbf{v}_h\mathbf{M}_r + \mathbf{r} - \mathbf{v}_t\mathbf{M}_r||_2^2 - ||\mathbf{v}_h\mathbf{M}_r + \mathbf{r} - \mathbf{v}_{t'}\mathbf{M}_r||_2^2) \\
& - \frac{\lambda_Z}{2}\sum_{l\notin\{\frac{L_v}{2},\frac{L_v}{2}+1\}} ||\sigma(\mathbf{Z}_{l-1} * \mathbf{Q}_l + \mathbf{c}_l) - \mathbf{Z}_l||_2^2 - \frac{\lambda_U}{2}\sum_i ||\mathbf{u}_i||_2^2 \\
& - \frac{\lambda_Z}{2}\sum_{l\in\{\frac{L_v}{2},\frac{L_v}{2}+1\}} ||\sigma(\mathbf{Z}_{l-1}\mathbf{Q}_l + \mathbf{c}_l) - \mathbf{Z}_l||_2^2 - \frac{\lambda_v}{2}\sum_v ||\mathbf{v}||_2^2 \\
& - \frac{1}{2}\sum_l(\lambda_W||\mathbf{W}_l||_2^2 + \lambda_b||\mathbf{b}_l||_2^2) - \frac{1}{2}\sum_l(\lambda_Q||\mathbf{Q}_l||_2^2 + \lambda_c||\mathbf{c}_l||_2^2) \\
& - \frac{\lambda_I}{2}\sum_j ||\mathbf{e}_j - \mathbf{v}_j - \mathbf{X}_{\frac{L_t}{2},j*} - \mathbf{Z}_{\frac{L_v}{2},j*}||_2^2 \\
& - \frac{\lambda_r}{2}\sum_r ||\mathbf{r}||_2^2 - \frac{\lambda_M}{2}\sum_r ||\mathbf{M}_r||_2^2
\end{aligned}
$$

联合训练

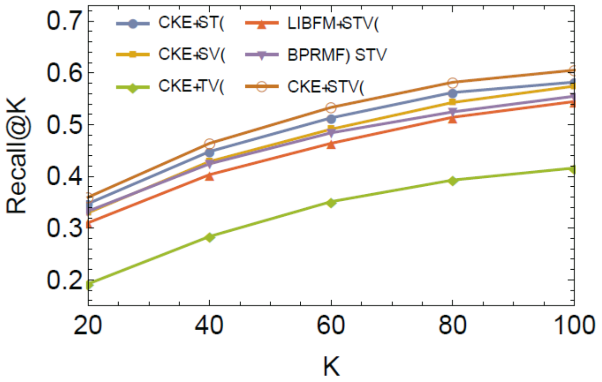
Collaborative Knowledge base Embedding (CKE)

性能对比

Recall@K

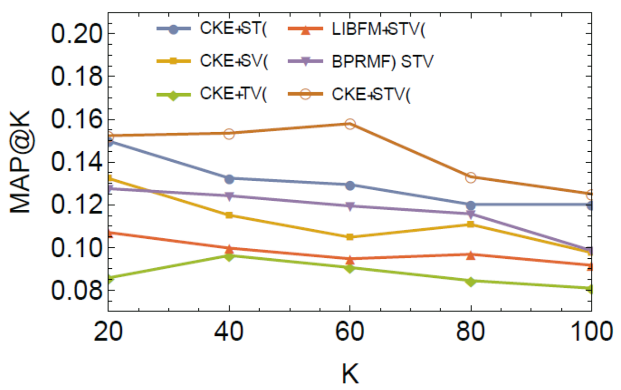


(a) MovieLens-1M

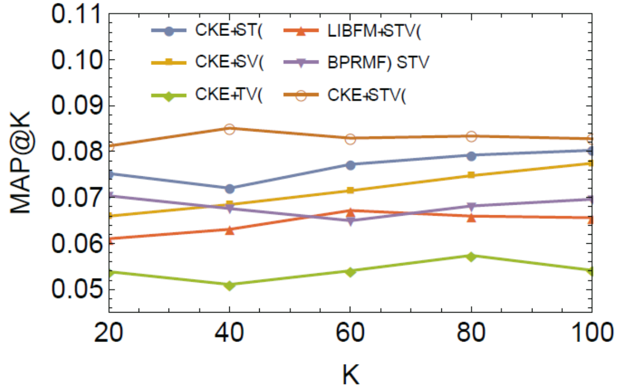


(b) IntentBooks

MAP@K



(a) MovieLens-1M



(b) IntentBooks

联合训练

Ripple Network

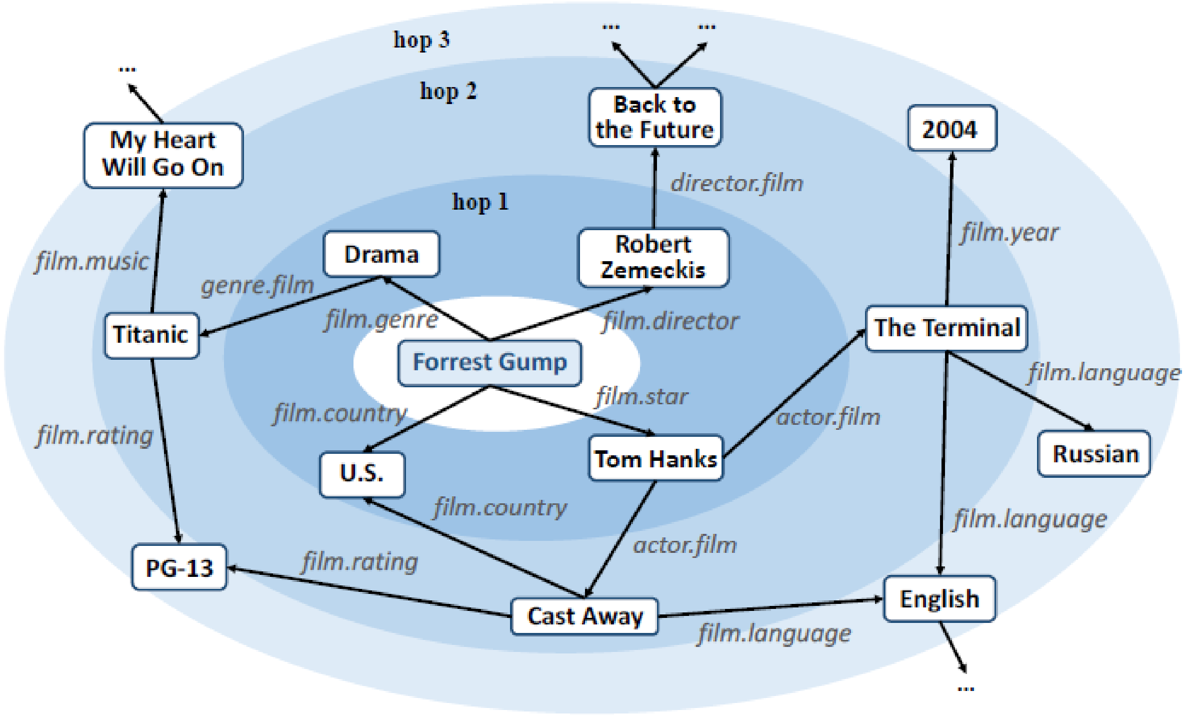
先前工作的不足：

- DKN，CKE等使用**知识图谱特征学习**作为关键步骤
  - 实体向量存在于隐含语义空间中，缺乏直观性
  - 现有的KGE方法更适合于连接预测等任务而非推荐
- MetaPath，MetaGraph等使用**自定义路径**来衡量实体间相似度
  - 需要手动设计metapath/metagraph，实践中难以达到最优
  - 在实体不属于同一个领域的场景中（如新闻推荐）无法应用

联合训练

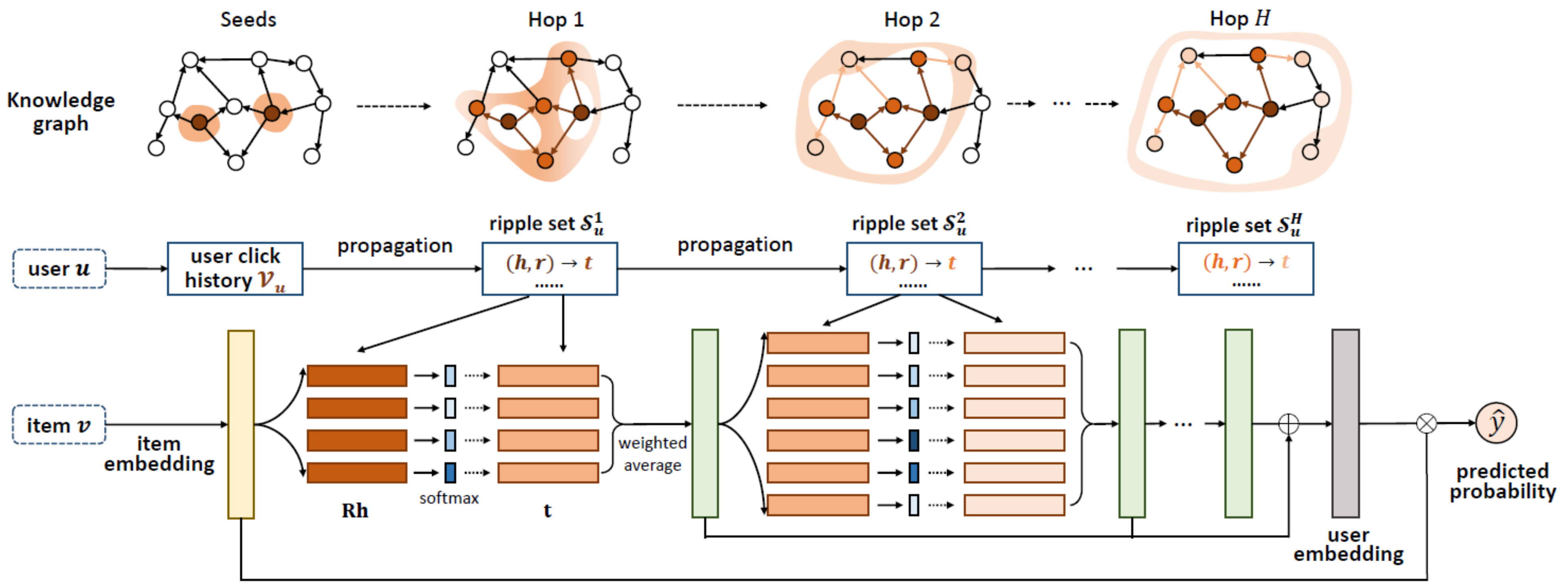
Ripple Network

- Ripple Network模拟用户兴趣在知识图谱上的传播过程
  - 用户兴趣以其历史记录为中心，在知识图谱上向外逐层扩散
  - 用户兴趣在扩散过程中逐渐衰减



联合训练

Ripple Network



$$\min \mathcal{L} = -\log\left(p(\mathrm{Y}|\Theta,\mathcal{G}) \cdot p(\mathcal{G}|\Theta) \cdot p(\Theta)\right)$$

$$= \sum_{(u,v)\in\mathrm{Y}} -\left(y_{uv}\log\sigma(\mathbf{u}^{\mathrm{T}}\mathbf{v}) + (1-y_{uv})\log\left(1-\sigma(\mathbf{u}^{\mathrm{T}}\mathbf{v})\right)\right)$$

$$+ \frac{\lambda_2}{2}\sum_{r\in\mathcal{R}}\|\mathbf{I}_r - \mathbf{E}^{\mathrm{T}}\mathbf{R}\mathbf{E}\|_2^2 + \frac{\lambda_1}{2}\left(\|\mathbf{V}\|_2^2 + \|\mathbf{E}\|_2^2 + \sum_{r\in\mathcal{R}}\|\mathbf{R}\|_2^2\right)$$

联合训练

Ripple Network

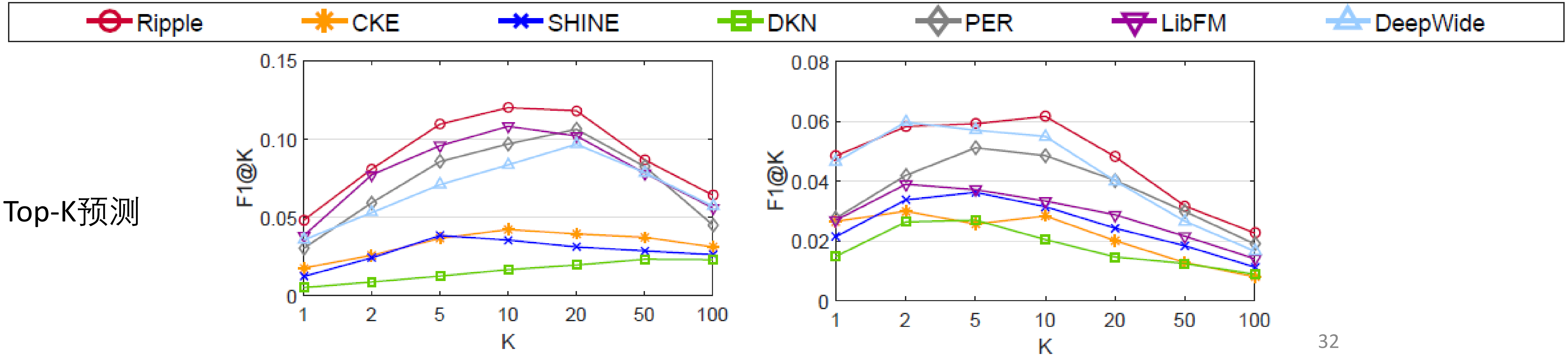
性能对比

CTR预测

Model	MovieLens-1M		Book-Crossing		Bing-News	
	AUC	*ACC*	*AUC*	*ACC*	*AUC*	*ACC*
Ripple*	**0.913**	**0.835**	**0.840**	**0.775**	**0.778**	**0.732**
CKE	0.796	0.739	0.634	0.606	0.660	0.617
SHINE	0.778	0.732	0.668	0.636	0.614	0.587
DKN	0.655	0.589	0.621	0.598	0.761	0.704
PER	0.901	0.826	0.814	0.735	-	-
LibFM	0.892	0.812	0.763	0.705	0.744	0.688
DeepWide	0.903	0.822	0.806	0.731	0.754	0.695

* Statistically significant improvements by t-test.

Top-K预测



MovieLens-1M

Book-Crossing

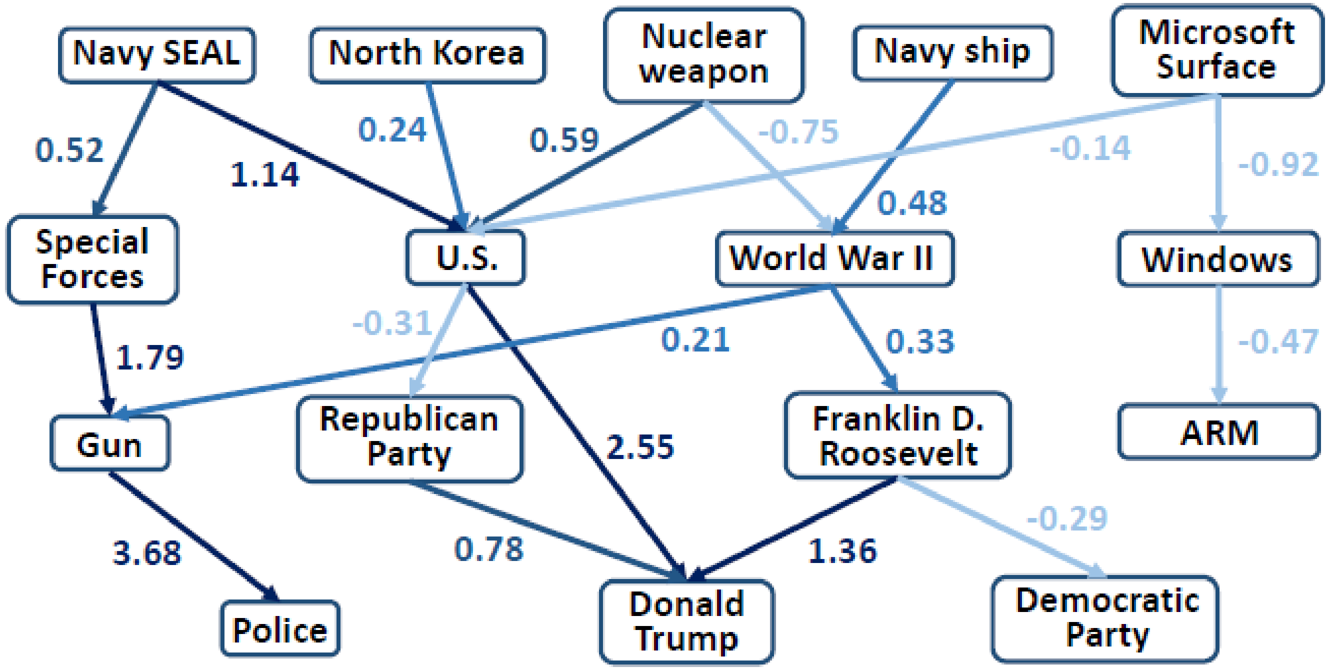
联合训练

Ripple Network

知识图谱用户兴趣传播的可视化

Click history:

1. *Family of **Navy SEAL** Trainee Who Died During Pool Exercise Plans to Take Legal Action*
2. ***North Korea** Vows to Strengthen **Nuclear Weapons***
3. ***North Korea** Threatens 'Toughest Counteraction' After **U.S.** Moves **Navy Ships***
4. *Consumer Reports Pulls Recommendation for **Microsoft Surface** Laptops*



Candidate news: *__Trump__ Announces Gunman Dead, Credits 'Heroic Actions' of __Police__*

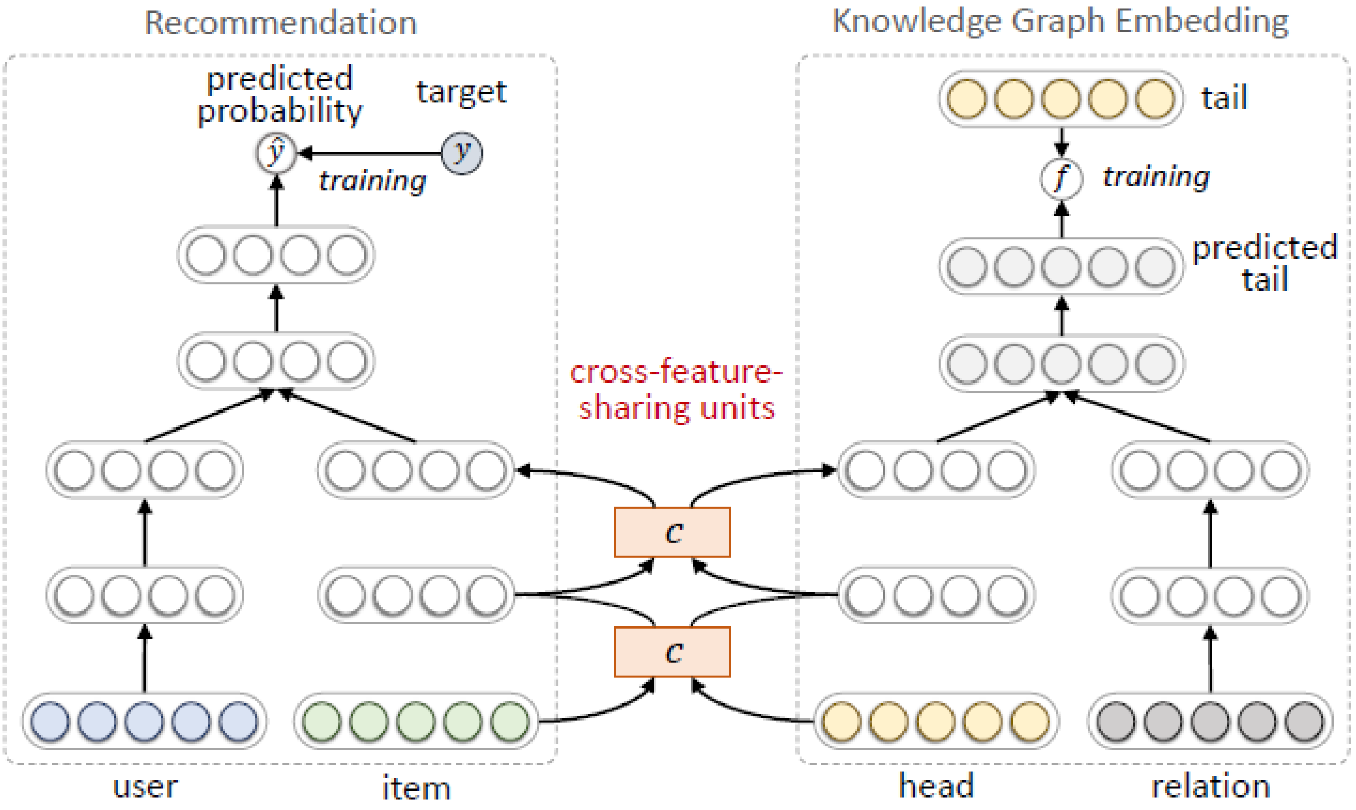
交替训练

Multi-task learning for Knowledge graph enhanced Recommendation (MKR)

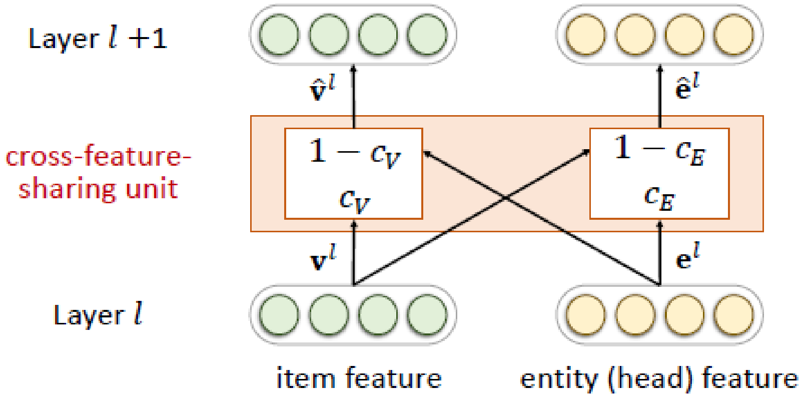
- 推荐系统中的物品和知识图谱中的实体存在重合
- 将推荐系统和KGE视为两个任务，采用多任务学习的框架
  - 两者的可用信息可以互补
  - KGE任务可以帮助推荐系统摆脱局部极小值
  - KGE任务可以防止推荐系统过拟合
  - KGE任务可以提高推荐系统的泛化能力

交替训练

Multi-task learning for Knowledge graph enhanced Recommendation (MKR)



MKR算法框架



cross-feature-sharing unit 结构

交替训练

Multi-task learning for Knowledge graph enhanced Recommendation (MKR)

Algorithm 1 Multi-Task Learning for MKR

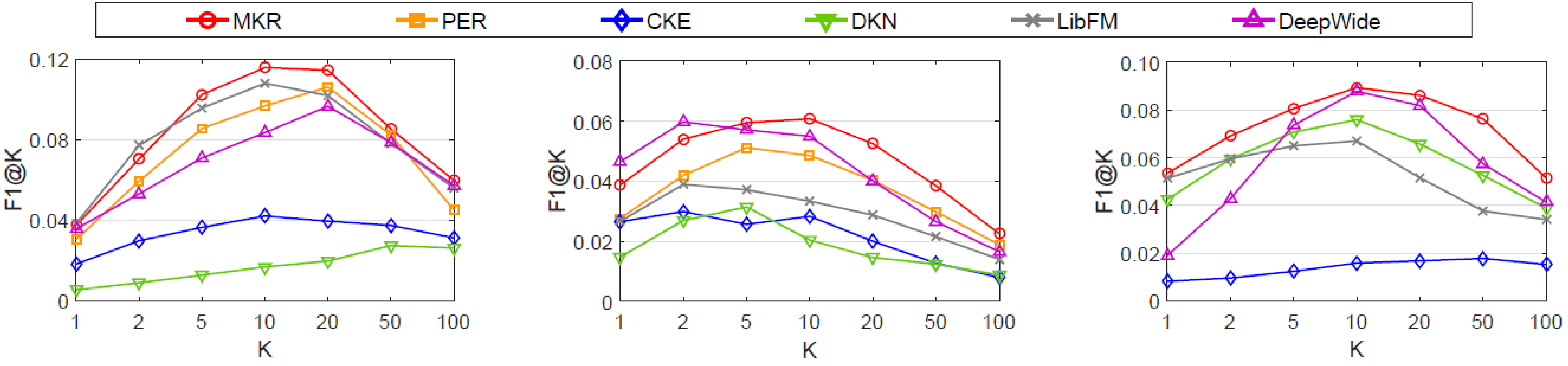
Input: Interaction matrix $\mathbf{Y}$, knowledge graph $\mathcal{G}$
Output: Prediction function $\mathcal{F}(u, v|\Theta, \mathbf{Y}, \mathcal{G})$

Initialize all parameters
for number of training iteration **do**
    // recommendation task
    for t steps **do**
        Sample minibatch of true and false interactions from $\mathbf{Y}$;
        Update parameters of $\mathcal{F}$ by SGD on Eq. (3)-(6), (9);
    end for
    // knowledge graph embedding task
    Sample minibatch of true and false triples from $\mathcal{G}$;
    Update parameters of $\mathcal{F}$ by SGD on Eq. (7)-(9);
end for

交替训练

Multi-task learning for Knowledge graph enhanced Recommendation (MKR)

性能对比



(a) MovieLens-1M (b) Book-Crossing (c) Bing-News

未来展望

- 知识图谱和时序模型的结合
  - 时序推荐（sequential recommendation）
  - 基于会话的推荐（session-based recommendation）
- 知识图谱与基于强化学习的推荐系统的结合
- 知识图谱与其它辅助信息在推荐系统中的有效结合
  - 社交网络（social network）
  - 用户属性（user attribute）
  - 物品属性（item attribute）
  - 上下文信息（context）


Question?