Inductive Representation Learning on Large Graphs

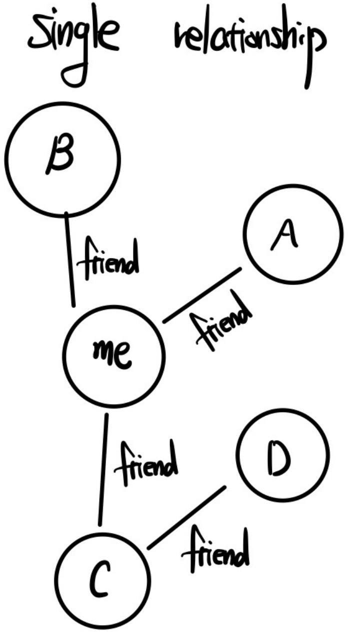
Published as a conference paper at NIPS 2017
William L. Hamilton, Rex Ying, Jure Leskovec

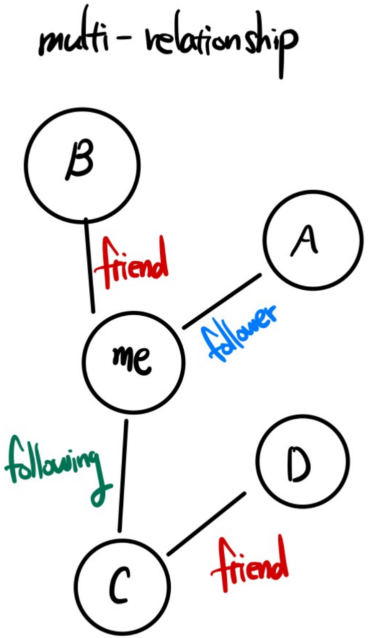
2024-07-21
HoonUi Lee

Contents

Previous work

In Social Network





Previous work

표현력, 보편성 증가를 목표

stochastic blockmodel

models based on tensor factorization

collective matrix factorization

→

모델 **복잡성** 증가

높은 **계산 비용**

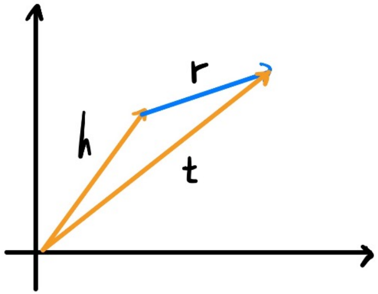
해석의 어려움

정규화 어려움

과적합 or **과소적합** 문제 발생

→ 성능은 비슷하게, 그러나 간단하게

TransE



$$\mathbf{h}+\mathbf{r} \approx \mathbf{t}$$

$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]$$

$$S'_{(h,\ell,t)} = \left\{(h',\ell,t)|h' \in E\right\} \cup \left\{(h,\ell,t')|t' \in E\right\}$$

Head와 Relation을 더했을 때, 최대한 Tail과 가깝도록

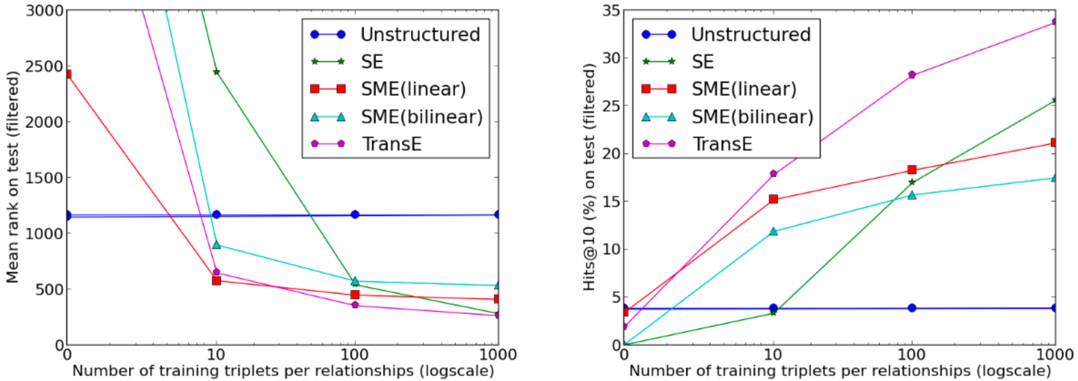
TransE

Mean rank: 모델이 제시한 정답 entity의 rank의 평균

Hits@10: 모델이 제시한 상위 10개의 정답 중 실제 정답이 있는 비율

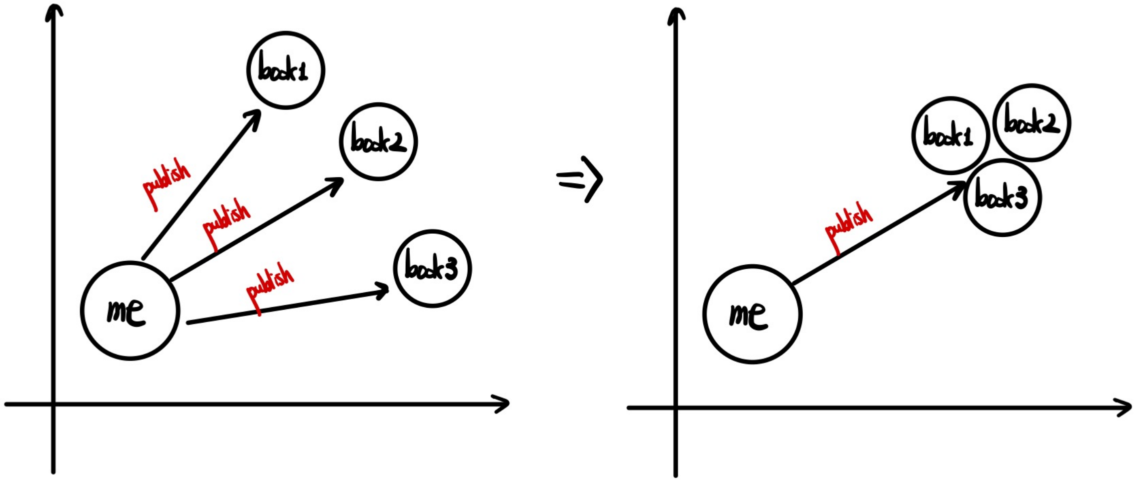
Table 1: **Numbers of parameters** and their values for FB15k (in millions). n_e and n_r are the nb. of entities and relationships; k the embeddings dimension.

METHOD	NB. OF PARAMETERS	ON FB15K
Unstructured [2]	$O(n_e k)$	0.75
RESCAL [11]	$O(n_e k + n_r k^2)$	87.80
SE [3]	$O(n_e k + 2n_r k^2)$	7.47
SME(LINEAR) [2]	$O(n_e k + n_r k + 4k^2)$	0.82
SME(BILINEAR) [2]	$O(n_e k + n_r k + 2k^3)$	1.06
LFM [6]	$O(n_e k + n_r k + 10k^2)$	0.84
TransE	$O(n_e k + n_r k)$	$\boxed{0.81}$



TransE

한계점: 1 - to - N



DistMult

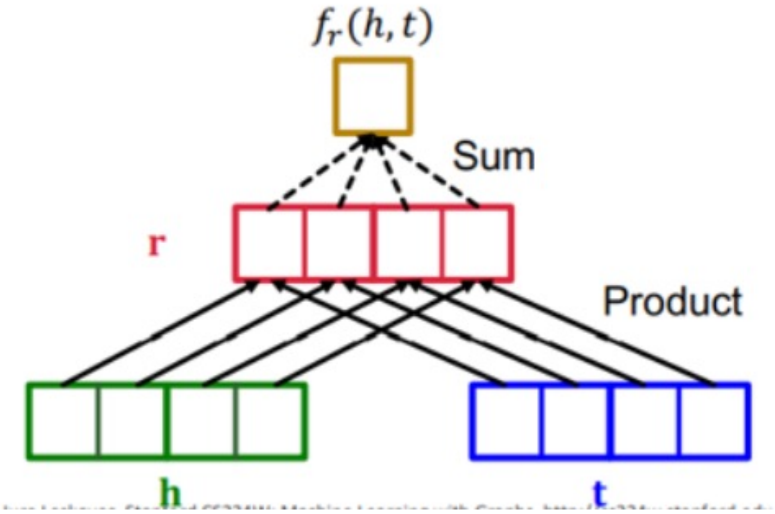
"모든 사람은 동물이다"

수량자 주어 서술어

자연어 문장에서의 Rule Extraction을

Knowledge Graph에서 시도

DistMult



DistMult

$$\mathbf{y}_{e_1} = f(\mathbf{W}\mathbf{x}_{e_1}), \quad \mathbf{y}_{e_2} = f(\mathbf{W}\mathbf{x}_{e_2})$$

$$g_r^a(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) = \mathbf{A}_r^T \begin{pmatrix} \mathbf{y}_{e_1} \\ \mathbf{y}_{e_2} \end{pmatrix}$$

Linear transformation

$$g_r^b(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) = \mathbf{y}_{e_1}^T \mathbf{B}_r \mathbf{y}_{e_2}$$

Bilinear transformation

Models	$\mathbf{B}_r$	$\mathbf{A}_r^T$	Scoring Function
Distance (Bordes et al., 2011)	-	$(\mathbf{Q}_{r1}^T \quad -\mathbf{Q}_{r2}^T)$	$-\lvert\lvert g_r^a(\mathbf{y}_{e_1}, \mathbf{y}_{e_2})\rvert\rvert_1$
Single Layer (Socher et al., 2013)	-	$(\mathbf{Q}_{r1}^T \quad \mathbf{Q}_{r2}^T)$	$\mathbf{u}_r^T \tanh(g_r^a(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}))$
TransE (Bordes et al., 2013b)	$\mathbf{I}$	$(\mathbf{V}_r^T \quad -\mathbf{V}_r^T)$	$-(2g_r^a(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) - 2g_r^b(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) + \lvert\lvert\mathbf{V}_r\rvert\rvert_2^2)$
NTN (Socher et al., 2013)	$\mathbf{T}_r$	$(\mathbf{Q}_{r1}^T \quad \mathbf{Q}_{r2}^T)$	$\boxed{\mathbf{u}_r^T \tanh\left(g_r^a(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) + g_r^b(\mathbf{y}_{e_1}, \mathbf{y}_{e_2})\right)}$

DistMult

$$g_r^b(\mathbf{y}_{e_1}, \mathbf{y}_{e_2}) = \mathbf{y}_{e_1}^T \boxed{\mathbf{M}_r} \mathbf{y}_{e_2}$$

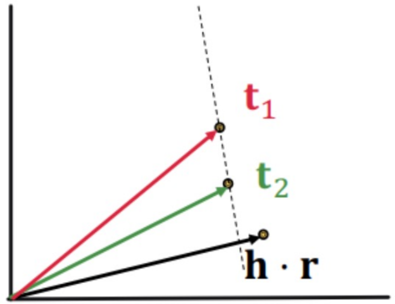
Diagonal matrix

r $\quad\quad$ M_r

$$[a.\ b.\ c] \Rightarrow \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

TransE와 parameter 수 동일

$$< \mathbf{h}, \mathbf{r}, \mathbf{t}_1 > = < \mathbf{h}, \mathbf{r}, \mathbf{t}_2 >$$



Experiment: Link Prediction

	FB15k		FB15k-401		WN	
	MRR	HITS@10	MRR	HITS@10	MRR	HITS@10
NTN	0.25	41.4	0.24	40.5	0.53	66.1
Blinear+Linear	0.30	49.0	0.30	49.4	0.87	91.6
TransE (DISTADD)	0.32	53.9	0.32	54.7	0.38	90.9
Bilinear	0.31	51.9	0.32	52.2	**0.89**	92.8
Bilinear-diag (DISTMULT)	**0.35**	**57.7**	**0.36**	**58.5**	0.83	**94.2**

MRR: Rank 역수의 평균

Hits@10: 모델이 제시한 상위 10개의 정답 중 실제 정답이 있는 비율

Experiment: Link Prediction

	Predicting subject entities				Predicting object entities			
	1-to-1	1-to-n	n-to-1	n-to-n	1-to-1	1-to-n	n-to-1	n-to-n
DISTADD	70.0	76.7	21.1	53.9	68.7	17.4	**83.2**	57.5
DISTMULT	**75.5**	**85.1**	**42.9**	**55.2**	**73.7**	**46.7**	81.0	**58.8**

Hits@10: 모델이 제시한 상위 10개의 정답 중 실제 정답이 있는 비율

Experiment: Link Prediction

선형? 비선형?

$$\mathbf{y}_{e_1} = \boxed{f}\left(\mathbf{W}\mathbf{x}_{e_1}\right), \quad \mathbf{y}_{e_2} = f\left(\boxed{\mathbf{W}}\mathbf{x}_{e_2}\right)$$

WV? EV? default?

	MRR	HITS@10	MAP (w/ type checking)
DISTMULT	0.36	58.5	64.5
DISTMULT-tanh	0.39	63.3	76.0
DISTMULT-tanh-WV-init	0.28	52.5	65.5
DISTMULT-tanh-EV-init	**0.42**	**73.2**	**88.2**

MAP: 모델이 예측한 rank list에서 실제 정답이 높은 rank에 위치 할수록 높은 수치

w/ type checking: relation에 맞는 entity type을 확인 (BornInCity(*Person*, *City*))

Experiment: Link Prediction

New / York / City

word vectors

New York City

entity vectors

Experiment: Rule Extraction

$$B_1(a_1, a_2) \wedge B_2(a_2, a_3) \wedge \ldots \wedge B_n(a_n, a_{n+1}) \implies H(a_1, a_{n+1})$$

Algorithm 1 EMBEDRULE

```
Input: KB = {(e_1, r, e_2)}, relation set R
Output: Candidate rules Q
for each r in R do
    Select the set of start relations S = {s : X_s ∩ X_r ≠ ∅}
    Select the set of end relations T = {t : Y_t ∩ Y_r ≠ ∅}
    Find all possible relation sequences
    Select the K-NN sequences P' ⊆ P for r based on dist(M_r, M_{p_1} ∘ ··· ∘ M_{p_n})
    Form candidate rules using P' where r is the head relation and p ∈ P' is the body in a rule
    Add the candidate rules into Q
end for
```

B1: 사람 a가 도시 b에서 태어났다.

B2: 도시 b는 국가 c에 속한다.

H: 사람 a의 국적은 국가 c이다.

Experiment: Rule Extraction

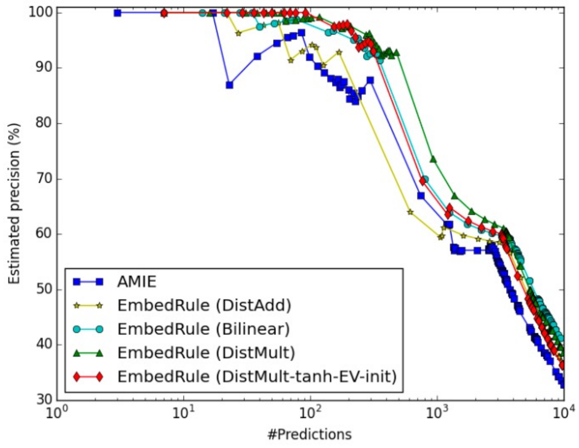


Figure 1: Aggregated precision of top length-2 rules extracted by different methods

Experiment: Rule Extraction

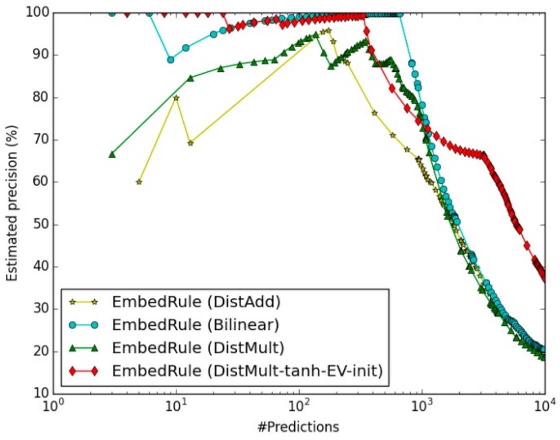
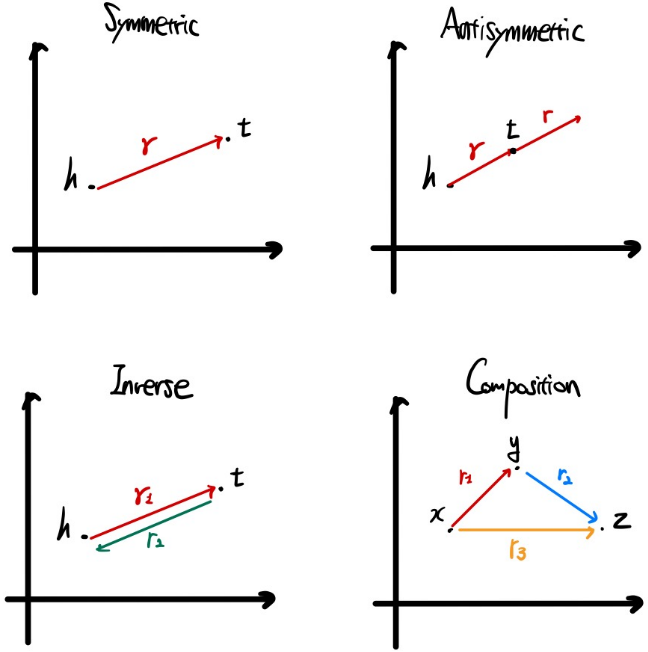


Figure 2: Aggregated precision of top length-3 rules extracted by different methods

Limitations: Relation Patterns

TransE



DistMult

Symmetric

$$\langle h, r, t \rangle = \langle t, r, h \rangle$$

Antisymmetric

$$\langle h, r, t \rangle = \langle t, r, h \rangle$$

Inverse

$$\langle h, r_1, t \rangle = \langle t, r_2, h \rangle$$

Composition

$$(M_{r_1} \cdot M_{r_2}) \neq M_{r_3}$$

Conclusion

Multi-relationship을 기존 연구보다 간단하게 표현한 TransE

내적을 활용하여 Multi-relationship 표현과 Rule Extraction 모두 두각을 드러낸 DistMult

Experiment를 통해 성능을 시연하고 다양한 확장을 도모하였지만, 여전히 한계점 존재