Translating Embeddings for Modeling Multi-relational Data

2022.07.19 김영로

Contents

Background(Word Embedding)

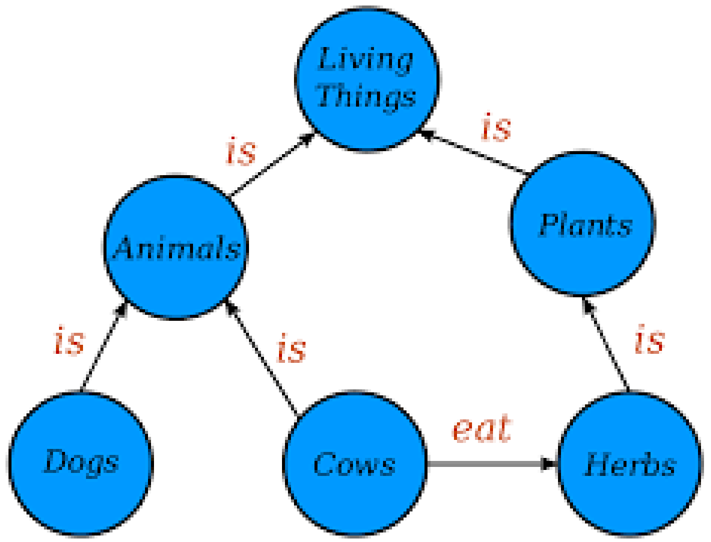
- 자연어(단어)를 벡터로 표현하는 방법
- one-hot vector 사용 시 공간적 낭비
- 차원을 줄여 사용하는 공간을 줄임

A 4-dimensional embedding

cat	=>	1.2	-0.1	4.3	3.2
mat	=>	0.4	2.5	-0.9	0.5
on	=>	2.1	0.3	0.1	0.4
...		...			

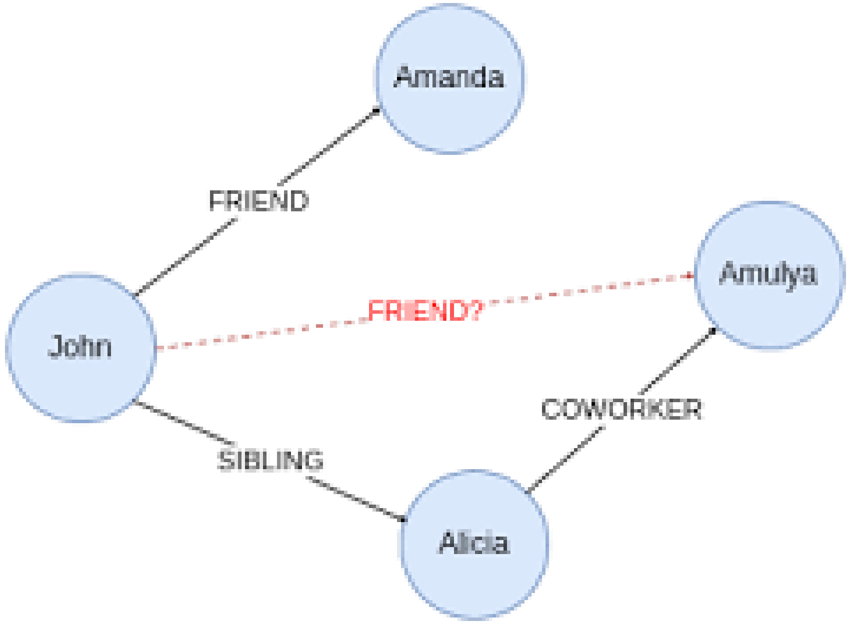
Background(Knowledge Graph)

- knowledge base: 어떤 분야에 관련된 지적 활동과 경험을 통해서 축적된 지식, 문제 해결에 필요한 사실과 규칙 등이 저장되어 있는 데이터베이스
- entity: 개체, relation: 개체 사이의 관계로 구성
- knowledge graph: 개체와 관계를 이용해서 생성되는 그래프
- entity를 node, relation을 edge로 표현함



Background(Knowledge Graph)

- embedding을 통해 변환된 데이터로 학습
- entity와 relation으로 학습된 모델을 이용해서 다음 entity 예측
- entity간의 relation 예측



Introduction

- node: entities, edge: relation

$(head, label, tail)$ denoted as (h, l, t)

- focuses on modeling multi-relational data from KBs

- Dataset: Wordnet, Freebase

- Goal: automatically adding new facts, without requiring extra knowledge

Introduction

Modeling multi-relational data

- ex) friend of my friend is my friend: relationship에 의존

- ex) Star Wars IV를 좋아하는 사람은 Star Wars V도 좋아할 가능성 높음, 하지만 Titanic에 대한 정보는 불확실: entity에 의존

- entities 사이의 pattern 도출이 목표

- entity와 relation을 동시에 고려할 수 있는 generic approaches가 필요함

Introduction

- 많은 recent approaches는 increasing the expressivity에 초점을 맞춤
- hard to interpret/ higher computational costs/ overfitting/ underfitting 과 같은 문제점들 발생 가능
- simple model can lead to better trade-offs between accuracy and scalability

Introduction

Relationships as translations in the embedding space

- TransE: energy-based model for learning low-dimensional embeddings of entities

- relationship은 translations in the embedding space에 표현됨

- (h, l, t) holds: t가 $h + l$에 가까움

Introduction

- main motivation: hierarchical relationships이 KBs에 자주 등장하고, translations이 이를 표현하는 natural transformation

- 다양한 type의 1-to-1 relationships between entities를 translation을 이용해서 표현하는 embedding space가 존재

- new model: simplicity, modeling hierarchies를 design한 architecture, outperform state-of-the-art methods in link prediction on real world KBs

Model

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\mathbf{e} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$S_{batch} \leftarrow$sample(S, b) // sample a minibatch of size b

$T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

for $(h, \ell, t) \in S_{batch}$ **do**

$(h', \ell, t') \leftarrow$sample$(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

end for

Update embeddings w.r.t. $\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$

end loop

Model

- $S = \{(h, l, t)\}$
- E: entities set, L: relationships set
- γ: margin
- k: embeddings dimension

Model

basic idea

- the functional relation induced by the l-labeled edges corresponds to a translation of the embeddings

- $\boldsymbol{h} + \boldsymbol{l} \approx \boldsymbol{t}$ when (h, l, t) holds ($\boldsymbol{t}$가 $\boldsymbol{h} + \boldsymbol{l}$이랑 가장 가까운 벡터)

- energy of a triplet = $d(\boldsymbol{h} + \boldsymbol{l}, \boldsymbol{t})$

- d: dissimilarity measure (L_1 or L_2-norm)

Model

Loss function

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

Set of corrupted triplets

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

주어진 entity에 대해 head나 tail에 나타날 때 embedding vector는 같음

Model

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$
        $\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$
        $\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$
loop
    $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$
    $S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b
    $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
    for $(h, \ell, t) \in S_{batch}$ **do**
        $(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet
        $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$
    end for
    Update embeddings w.r.t. $\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla [\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})]_+$
end loop

initialize $\boldsymbol{l}, \boldsymbol{e}$
normalize $\boldsymbol{l}$

for each loop

- normalize $\boldsymbol{e}$
- sample minibatch of size b
- initialize T_{batch}
- for each element in S_{batch}, sample corruted triplet and add to T_{batch}
- Update embeddings

Experiment

Data sets(Wordnet)

- entities correspond to word senses, relationships define lexical relations between them
- denote as WN


train.txt - 메모장

파일 편집 보기

03964744	04371774	_hyponym
00260881	00260622	_hypernym
02199712	02188065	_member_holonym
01332730	03122748	_derivationally_related_form
06066555	00645415	_derivationally_related_form
09322930	09360122	_instance_hypernym
11575425	12255934	_hyponym
07193596	00784342	_derivationally_related_form
05726596	06162979	_hyponym
01768969	02636811	_derivationally_related_form
02557199	02557790	_hyponym
01455754	01974062	_hypernym
02716866	03032576	_hyponym
03214670	04423288	_hyponym
07554856	07553301	_hypernym
11669921	11992806	_hyponym
01291069	01530678	_hyponym
07965085	08278169	_hyponym
00057306	00056912	_hypernym
10341660	02661252	_derivationally_related_form
13219258	13167078	_hypernym
01698271	01754576	_also_see
08189659	08077292	_hyponym
10499355	10083823	_hypernym
02222318	02223238	_hyponym
02103406	02084071	_hypernym
07190941	07185325	_hypernym
12090318	12093769	_member_meronym
08620061	08620763	_hyponym
03562126	03318438	_hyponym
12213635	12214245	_member_meronym
02651424	02672371	_derivationally_related_form
13278375	13282007	_hyponym
06090869	02835887	_member_of_domain_topic
10668450	10525134	_hypernym
01837746	01342529	_hypernym
04692908	01259005	_derivationally_related_form
00848169	06719579	_derivationally_related_form
10483138	00968211	_derivationally_related_form
03631445	01521603	_derivationally_related_form

Experiment

Data sets(Freebase)

- around 1.2 billion triplets, 80 million entities about general facts
- two data set from Freebase
  - 1. FB15k: Wikilinks database에 나오며 100번 이상 언급된 entities
  - 2. FB1M: 가장 많이 언급된 entities 1,000,000개


train.txt - 메모장

파일 편집 보기

```
/m/027rn	/m/06cx9	/location/country/form_of_government
/m/017dcd	/m/06v8s0	/tv/tv_program/regular_cast./tv/regular_tv_appearanc
/m/07s9rl0	/m/0170z3	/media_common/netflix_genre/titles
/m/01sl1q	/m/044mz_	/award/award_winner/awards_won./award/award_honor/aw
/m/0cnk2q	/m/02nzb8	/soccer/football_team/current_roster./sports/sports_
/m/02_j1w	/m/01cwm1	/sports/sports_position/players./soccer/football_ros
/m/059ts	/m/03h_f4	/government/political_district/representatives./gove
/m/011yn5	/m/01pjr7	/film/film/starring./film/performance/actor
/m/04nrcg	/m/02sdk9v	/soccer/football_team/current_roster./soccer/footbal
/m/07nznf	/m/014lc_	/film/actor/film./film/performance/film
/m/05cvgl	/m/04kxsb	/award/award_nominated_work/award_nominations./awarc
/m/02qyp19	/m/02d413	/award/award_category/nominees./award/award_nominati
/m/02vk52z	/m/01crd5	/organization/membership_organization/members./organ
/m/0q9kd	/m/0184jc	/award/award_nominee/award_nominations./award/award_
/m/09w1n	/m/0sx8l	/olympics/olympic_sport/olympic_games_contested
/m/03q5t	/m/07y_7	/music/performance_role/regular_performances./music/
/m/0gqng	/m/073hkh	/award/award_category/winners./award/award_honor/cer
/m/0b76d_m	/m/029j_	/film/film/release_date_s./film/film_regional_releas
/m/04ztj	/m/012v9y	/people/marriage_union_type/unions_of_this_type./peo
/m/014_x2	/m/012ljv	/award/award_winning_work/awards_won./award/award_ho
/m/0ds35l9	/m/05r4w	/film/film/release_date_s./film/film_regional_releas
/m/015qsq	/m/02bjrlw	/film/film/language
/m/04bdxl	/m/02s2ft	/award/award_nominee/award_nominations./award/award_
/m/0677ng	/m/03ckfl9	/music/artist/genre
/m/03zj9	/m/059rby	/location/location/containedby
/m/0157m	/m/063k3h	/people/person/ethnicity
/m/01ky7c	/m/019v9k	/education/educational_institution/students_graduate
/m/02681vq	/m/01w61th	/award/award_category/nominees./award/award_nominati
/m/0gx_st	/m/09qvf4	/award/award_ceremony/awards_presented./award/award_
/m/09c7w0	/m/0rs6x	/location/location/contains
/m/079vf	/m/0d90m	/film/actor/film./film/performance/film
/m/015zyd	/m/09c7w0	/organization/organization/headquarters./location/ma
/m/05vsxz	/m/06qgvf	/award/award_nominee/award_nominations./award/award_
/m/04ljl_l	/m/03qcfvw	/award/award_category/nominees./award/award_nominati
/m/0grwj	/m/05sxg2	/people/person/profession
/m/02ynfr	/m/0bmch_x	/film/film_job/films_with_this_crew_job./film/film_c
/m/018gz8	/m/071wvh	/people/profession/people_with_this_profession
/m/02cqbx	/m/01jr4j	/film/film_costumer_designer/costume_design_for_film
/m/0rh6k	/m/01lp8	/location/statistical_region/religions./location/rel
/m/03mqtr	/m/04j4tx	/film/film_genre/films_in_this_genre
```

Table 2: **Statistics of the data sets used in this paper and extracted from the two knowledge bases, Wordnet and Freebase.**

DATA SET	WN	FB15K	FB1M
ENTITIES	40,943	14,951	1×10^6
RELATIONSHIPS	18	1,345	23,382
TRAIN. EX.	141,442	483,142	17.5×10^6
VALID EX.	5,000	50,000	50,000
TEST EX.	5,000	59,071	177,404

Experiment

Setup (Evaluation protocol)

- ranking procedure
- corrupted triplets의 dissimilarities가 model에 의해 계산된 후, ascending order로 sort됨 (head, tail 두 경우) 그 후 correct entity의 rank를 저장
  - mean of predicted rank
  - hits@10 (상위 10위에 올바른 entities의 비율)
- To avoid error, use filtered rank

Experiment

learning rate λ: {0.001, 0.01, 0.1}
margin γ: {1, 2, 10}
dimension k: 20, 50
dissimilarity measure d: L_1, L_2
epochs: 1,000 (early stopping)

Optimal configurations (λ, γ, k, d)
WN: (0.01, 2, 20, L_1)
FB15k: (0.01, 1, 50, L_1)
FB1M: (0.01, 1, 50, L_2)

Related work (method)

- Unstructured: consider data as mono-relational and sets all translations to $\mathbf{0}$
- RESCAL: collective matrix factorization model
- SE: embeds entities into $\mathbb{R}^k$, relationship into two matrices $L_1, L_2 \in \mathbb{R}^{k \times k}$
- SE, SME(linear), SME(bilinear), LFM: energy-based model

Experiment

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	$O(n_e k + n_r k)$	0.81

Table 1 compares the theoretical number of parameters of baseline of each model

In case of SE, RESCAL, they learn at least one $k \times k$ matrix for each relationship so they need many parameters than others

Experiment

Table 3: **Link prediction results.** Test performance of the different methods.

DATASET	WN				FB15K				FB1M	
METRIC	MEAN RANK		HITS@10 (%)		MEAN RANK		HITS@10 (%)		MEAN RANK	HITS@10 (%)
Eval. setting	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Raw*
Unstructured [2]	315	304	35.3	38.2	1,074	979	4.5	6.3	15,139	2.9
RESCAL [11]	1,180	1,163	37.2	52.8	828	683	28.4	44.1	-	-
SE [3]	1,011	985	68.5	80.5	273	162	28.8	39.8	22,044	17.5
SME(LINEAR) [2]	545	533	65.1	74.1	274	154	30.7	40.8	-	-
SME(BILINEAR) [2]	526	509	54.7	61.3	284	158	31.3	41.3	-	-
LFM [6]	469	456	71.4	81.6	283	164	26.0	33.1	-	-
TransE	**263**	**251**	**75.4**	**89.2**	**243**	**125**	**34.9**	**47.1**	**14,615**	**34.0**

- filtered setting provides lower mean ranks, higher hits@10
- generally, trends between raw and filtered are the same

Experiment

- 저자는 TransE의 good performance가 appropriate design of the model, its relative simplicity로부터 나온다고 생각
- stochastic gradient를 이용한 효율적인 최적화 가능
- translation term의 impact가 큼

Experiment

Issue of other model

- SE: more expressive, complexity make it hard to learn

- SME, LFM: not enough train so could not exploit their full capabilities

- Unstructured: simply clusters all entities co-occurring together, independent to relationship, make guess only which entities are related

Experiment

Table 4: **Detailed results by category of relationship.** We compare Hits@10 (in %) on FB15k in the filtered evaluation setting for our model, TransE and baselines. (M. stands for MANY).

TASK	PREDICTING *head*				PREDICTING *tail*			
REL. CATEGORY	1-TO-1	1-TO-M.	M.-TO-1	M.-TO-M.	1-TO-1	1-TO-M.	M.-TO-1	M.-TO-M.
Unstructured [2]	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE [3]	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME(LINEAR) [2]	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
SME(BILINEAR) [2]	30.9	**69.6**	**19.9**	38.6	28.2	13.1	**76.0**	41.8
TransE	**43.7**	65.7	18.2	**47.2**	**43.7**	**19.7**	66.7	**50.0**

- categorized the relationships: 1-To-1, 1-To-M, M-To-1, M-To-M
- average number (appear in dataset)>1.5 -> Many

Experiment

Table 5: **Example predictions** on the FB15k test set using TransE. **Bold** indicates the test triplet's true tail and *italics* other true tails present in the training set.

INPUT (HEAD AND LABEL)	PREDICTED TAILS
J. K. Rowling `influenced by`	*G. K. Chesterton*, J. R. R. Tolkien, *C. S. Lewis*, **Lloyd Alexander**, Terry Pratchett, Roald Dahl, Jorge Luis Borges, *Stephen King*, Ian Fleming
Anthony LaPaglia `performed in`	*Lantana*, *Summer of Sam*, *Happy Feet*, *The House of Mirth*, Unfaithful, **Legend of the Guardians**, Naked Lunch, X-Men, The Namesake
Camden County `adjoins`	**Burlington County**, *Atlantic County*, *Gloucester County*, Union County, Essex County, New Jersey, Passaic County, Ocean County, Bucks County
The 40-Year-Old Virgin `nominated for`	*MTV Movie Award for Best Comedic Performance*, *BFCA Critics' Choice Award for Best Comedy*, *MTV Movie Award for Best On-Screen Duo*, MTV Movie Award for Best Breakthrough Performance, **MTV Movie Award for Best Movie**, MTV Movie Award for Best Kiss, D. F. Zanuck Producer of the Year Award in Theatrical Motion Pictures, Screen Actors Guild Award for Best Actor - Motion Picture
Costa Rica football team `has position`	*Forward*, *Defender*, *Midfielder*, **Goalkeepers**, Pitchers, Infielder, Outfielder, Center, Defenseman
Lil Wayne `born in`	**New Orleans**, Atlanta, Austin, St. Louis, Toronto, New York City, Wellington, Dallas, Puerto Rico
WALL-E `has the genre`	Animations, Computer Animation, *Comedy film*, *Adventure film*, *Science Fiction*, **Fantasy**, Stop motion, *Satire*, Drama

- input: head and label -> output: top predicted tails and true one
- good answer가 항상 top rank에 있지는 않아도, prediction은 common sense를 보임

Experiment

predict new relationships (setup)

- select 40 random relationships
- split data into two set: FB15k-40rel, FB15k-rest
- FB15k-rest: 353,788 triplets training set, 53,266 triplets validation set
- FB15k-40rel: 40,000 training set, 45,159 test set

Experiment

process

1. model을 FB15k-rest를 이용해 train, select with validation set
2. FB15k-40rel을 이용해 40 relations에 연관된 parameter를 learn
3. FB15k-40rel의 test set을 이용해 evaluate
4. repeat this procedure using 0, 10, 100, and 1000 examples of each relationship in phase 2

Experiment

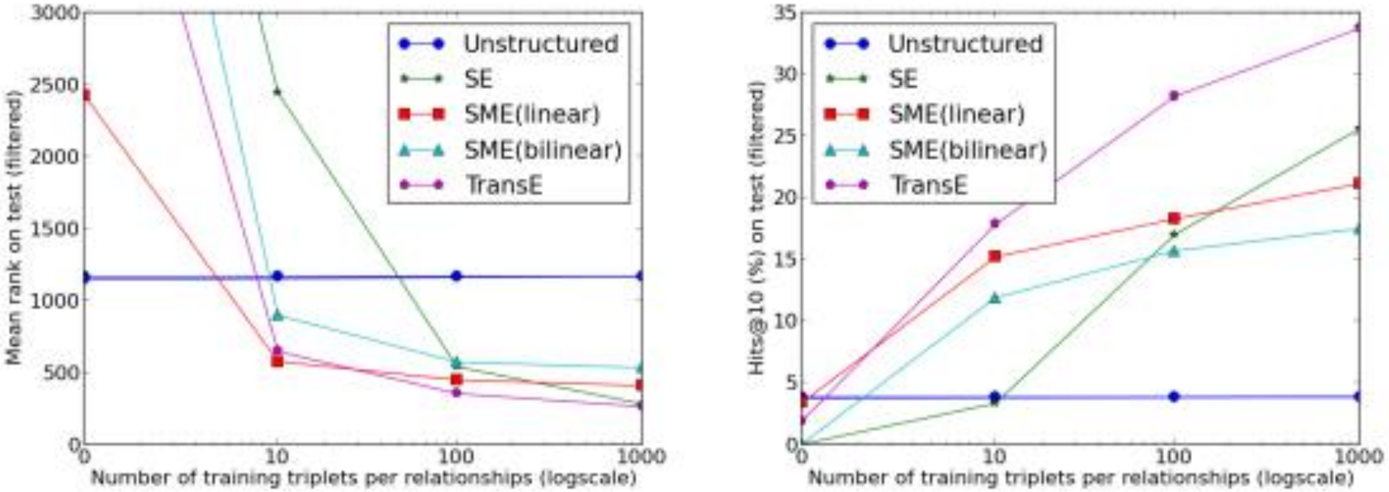


Figure 1: **Learning new relationships with few examples**. Comparative experiments on FB15k data evaluated in mean rank (left) and hits@10 (right). More details in the text.

simplicity of the TransE model makes it able to generalize well

Conclusion

- Find new approach to learn embeddings of KBs
- minimal parametrization to represent hierarchical relationships
- works very well compared with other methods on two data base
- highly scalable model
- all relationship types can be modeled adequately by this approach

Thank you!