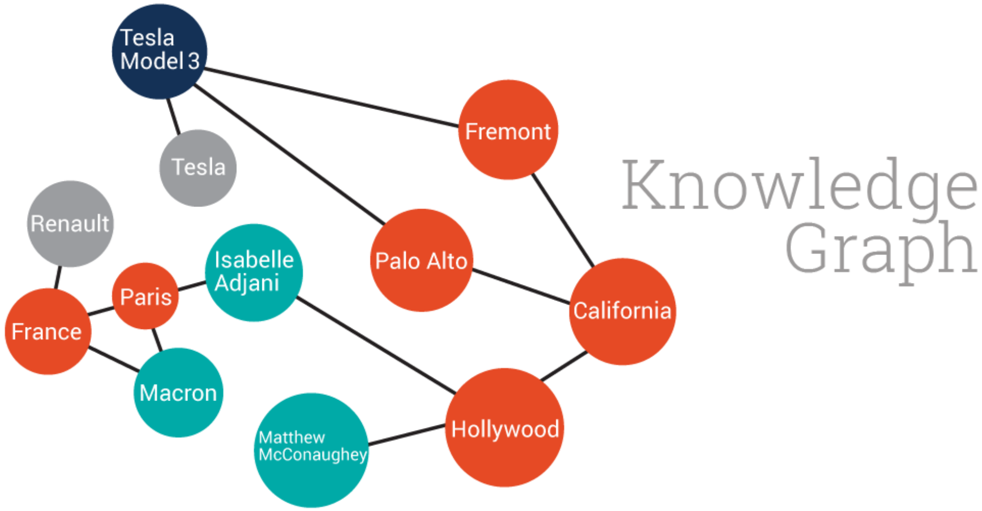
Learning entity and relation embeddings for knowledge graph completion

Review

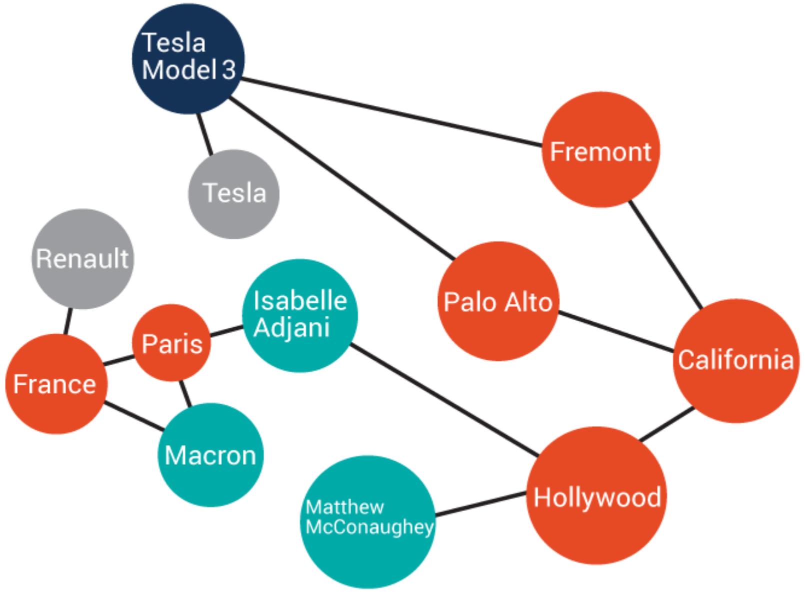
Learning entity and relation embeddings for knowledge graph completion

Knowledge Graph completion



"Aim to predict relations between entities under supervision of the existing knowledge graph"

01 Introduction
Knowledge Graph completion Challenge



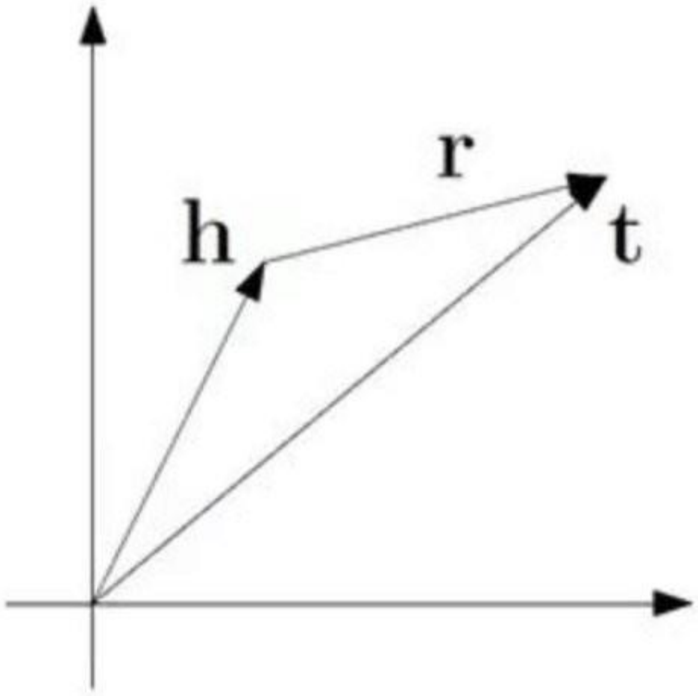
1. Nodes in knowledge graphs are entities with different types and attributes
2. Edges in knowledge graphs are relations of different types

-> link prediction is not capable

so, TransE and TransH used

02 Related Models

TransE



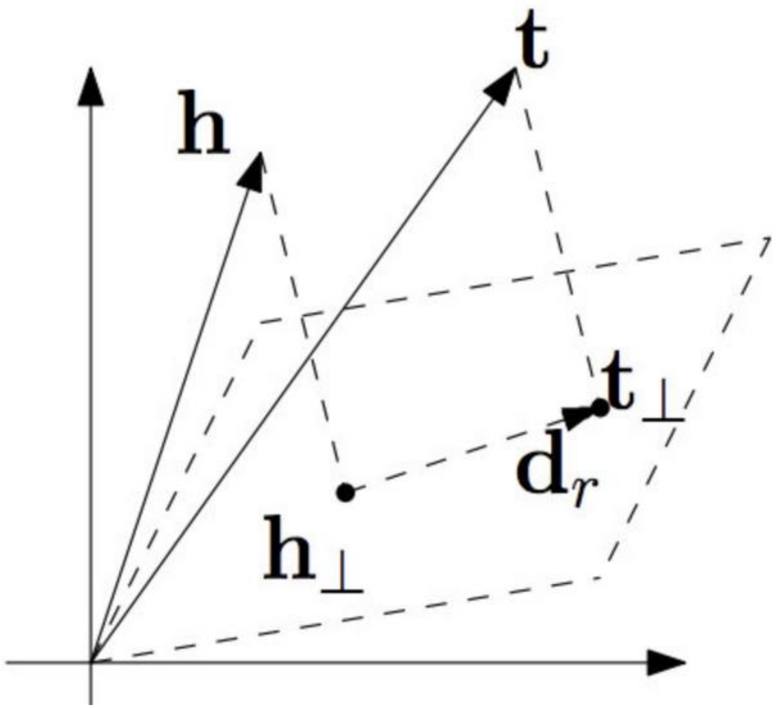
$$f_r(h, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2$$

- Head + relation = tail
- Score function has a purpose

  that head+ relation node is similar to tail node

Issue of TransE : N-to-1, 1-to-N, N-to-N relation

02 Related Models
TransH



$$\|\mathbf{h}_\perp + \mathbf{d}_r - \mathbf{t}_\perp\|_2^2$$

Score function of TransH

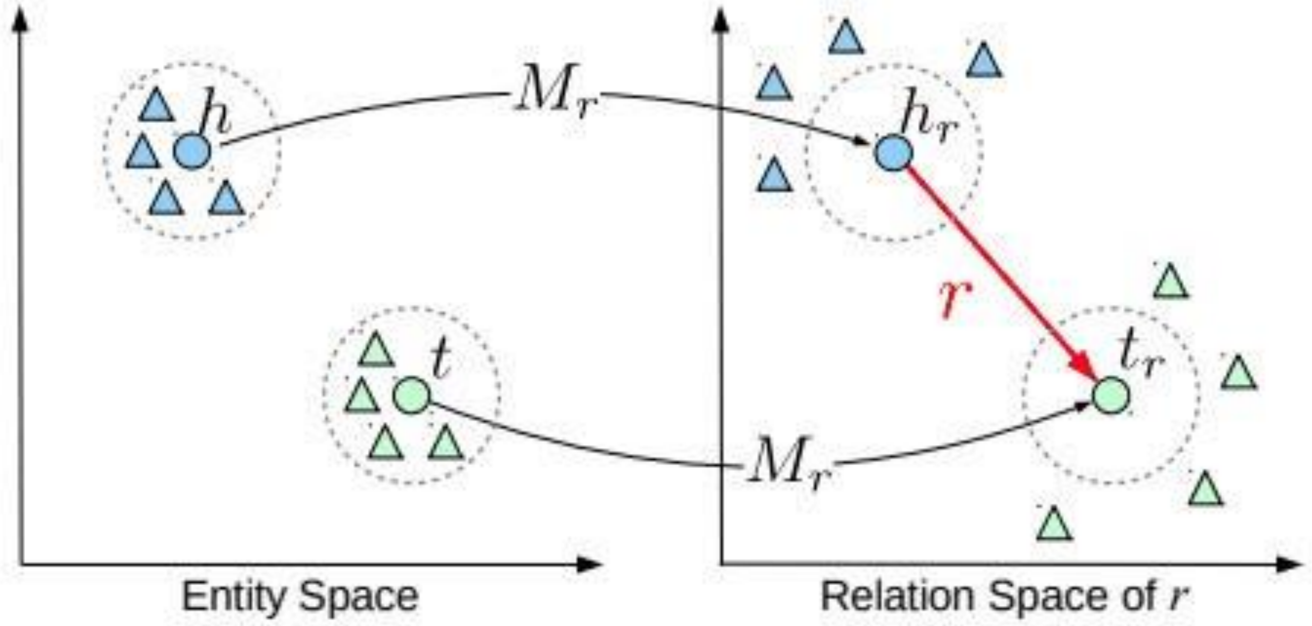
$$\mathbf{h}_\perp = \mathbf{h} - \mathbf{w}_r^\top \mathbf{h}\mathbf{w}_r, \quad \mathbf{t}_\perp = \mathbf{t} - \mathbf{w}_r^\top \mathbf{t}\mathbf{w}_r.$$

- Project the relation into hyperplane
- Every node can be distinguished from each relation

But, Entities and relations are in the same embedding space

02 Related Models

TransR



$$\mathbf{h}_r = \mathbf{h}\mathbf{M}_r, \quad \mathbf{t}_r = \mathbf{t}\mathbf{M}_r.$$

$$f_r(h, t) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_2^2.$$

Entities and relations in distinct spaces and performs translation in relation space

02 Related Models
CTransR : Cluster-based TransR

Incorporate the idea of piecewise linear regression to extend TransR

-> better model (TransE, TransH, TransR) relations

1. Segment input instance into several groups
2. Clustered into multiple groups

$$f_r(h,t) = \|\mathbf{h}_{r,c} + \mathbf{r}_c - \mathbf{t}_{r,c}\|_2^2 + \alpha\|\mathbf{r}_c - \mathbf{r}\|_2^2$$

<Score function: Not too far away from the original relation vector r>

03 Training method and implementation details

S = the set of correct triples / S' is the set of incorrect triple

$$L = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \max\left(0, \underbrace{f_r(h,t)}_{\text{positive}} + \underbrace{\gamma}_{\text{margin}} - \boxed{\underbrace{f_r(h',t')}_{\text{negative : tail}}}\right),$$

If A == 0
Ideal result

< Margin-based score function >

Sampling method (unif, bern)

(bern : reducing false negative labels by replacing head or tail with different probabilities)

=> Use stochastic gradient descent (확률적 경사 하강법)

Avoid Over-fitting

- Initialize entity and relation embeddings with result of TransE
- Initialize relation matrices as identity matrices (단위 행렬)

04 Data Set

WordNet (WN18, WN11) : Each entity is a synset consisting of several words

Freebase (FB15K) : General facts of the world

Table 1: Statistics of data sets.

Dataset	#Rel	#Ent	#Train	#Valid	# Test
WN18	18	40,943	141,442	5,000	5,000
FB15K	1,345	14,951	483,142	50,000	59,071
WN11	11	38,696	112,581	2,609	10,544
FB13	13	75,043	316,232	5,908	23,733
FB40K	1,336	39528	370,648	67,946	96,678

05 Experiment
1) Link prediction

Evaluation measure

- Mean rank of correct entities
- Proportion of correct entities in top-10 (Hits@10)

Table 2: Evaluation results on link prediction.

Data Sets	WN18				FB15K			
Metric	Mean Rank		Hits@10 (%)		Mean Rank		Hits@10 (%)	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickel, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) (Bordes et al. 2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif) (Wang et al. 2014)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern) (Wang et al. 2014)	401	388	73.0	82.3	212	87	45.7	64.4
TransR (unif)	232	219	78.3	91.7	226	78	43.8	65.5
TransR (bern)	238	225	**79.8**	92.0	**198**	77	48.2	68.7
CTransR (unif)	243	230	78.9	**92.3**	233	82	44	66.3
CTransR (bern)	**231**	**218**	79.4	**92.3**	199	**75**	**48.4**	**70.2**

①

②

③

Evaluation Result

1) TransR and CTransR >> TransE and TransH

2) CTransR >> TransR

3) Bern Sampling tricks works better

Separate evaluation result

- Predicting 1-to-1 relation -> more precise representation
- Predicting 1-to-N, N-to-1 relation -> discriminate relevant from irrelevant entities

Table 3: Evaluation results on FB15K by mapping properties of relations. (%)

Tasks	Predicting Head(Hits@10)				Predicting 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
Unstructured (Bordes et al. 2012)	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE (Bordes et al. 2011)	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME (linear) (Bordes et al. 2012)	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
SME (bilinear) (Bordes et al. 2012)	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE (Bordes et al. 2013)	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH (unif) (Wang et al. 2014)	66.7	① 81.7	30.2	57.4	63.7	30.1	83.2	60.8
TransH (bern) (Wang et al. 2014)	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransR (unif)	76.9	77.9	**38.1**	66.9	76.2	38.4	76.2	69.1
TransR (bern)	78.8	**89.2**	34.1	69.2	79.2	37.4	**90.4**	72.1
CTransR (unif)	78.6	77.8	36.4	68.0	77.4	37.8	78.0	70.3
CTransR (bern)	**81.5**	89.0	34.7	**71.2**	**80.8**	38.6	90.1 ②	**73.8**

05 Experiment
2) Triple Classification

Judge whether (h, r, t) is correct or not -> need negative triple : Use NTN

Table 5: Evaluation results of triple classification. (%)

Data Sets	WN11	FB13	FB15K
SE	53.0	75.2	-
SME (bilinear)	70.0	63.7	-
SLM	69.9	85.3	-
LFM	73.8	84.3	-
NTN	70.4	**87.1** ②	68.5
TransE (unif)	75.9	70.9	79.6
TransE (bern)	75.9	81.5	79.2
TransH (unif)	77.7	76.5	79.0
TransH (bern)	78.8	83.3	80.2
TransR (unif)	85.5	74.7	81.7 ③
TransR (bern)	① **85.9**	82.5	83.9
CTransR (bern)	85.7	-	**84.5**

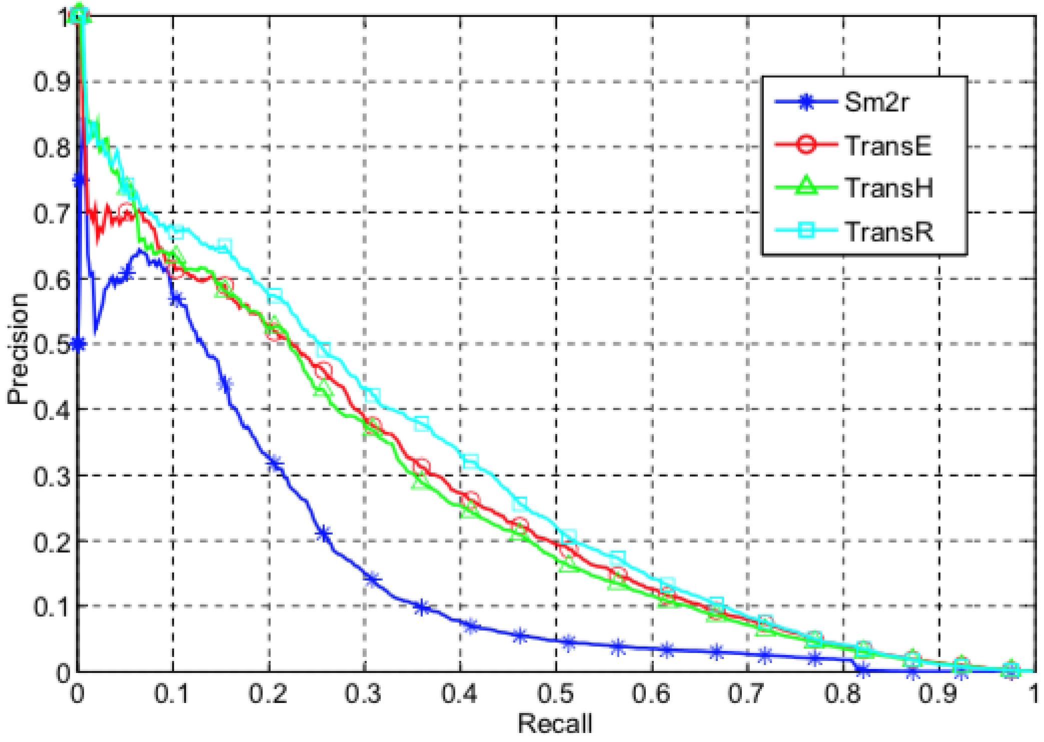
Triple Classification Result

1) WN11 : TransR >> TransE and TransH
2) FB13 : NTN >> TransE, TransH, TransR
   (NTN is the most expressive model)
3) Bern Sampling tricks works better

05 Experiment
3) Relation Extraction from text

Principle : Extract relational fact
from large-scale plain text

Result : TransR has better precision



06 Conclusion

- TransR embeds entities and relations
  in distinct entity space and relation space

- CTransR aims to model internal complicated correlations
  within each relation type

=> TransR achieves consistent and significant improvements
compared to TransE and TransH

THANK YOU