Trans* and DeepWalk

by Ran Jia (22 Apr 2016)
and Lili Mou (additional reading)

Trans*

by Ran Jia

2016.4.22

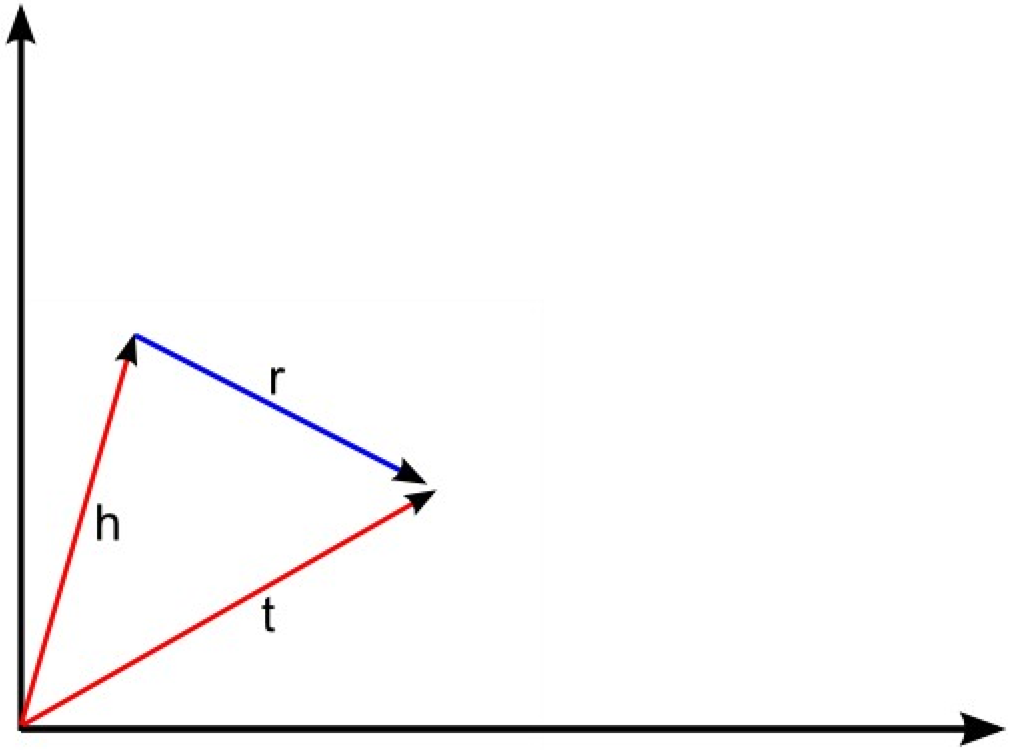
Relational fact: (head, relation, tail)

TransE:
Translating embeddings for modeling multi-relational data. Bordes, Antoine, et al. *Advances in Neural Information Processing Systems*. 2013.

TransH:
Knowledge Graph Embedding by Translating on Hyperplanes. Wang, Zhen, et al. *AAAI*. 2014.

TransR :
Learning Entity and Relation Embeddings for Knowledge Graph Completion. Lin, Yankai, et al. *AAAI*. 2015.

TransE



◆$h + r \approx t$ when $(\mathrm{h}, \mathrm{r}, \mathrm{t})$ holds

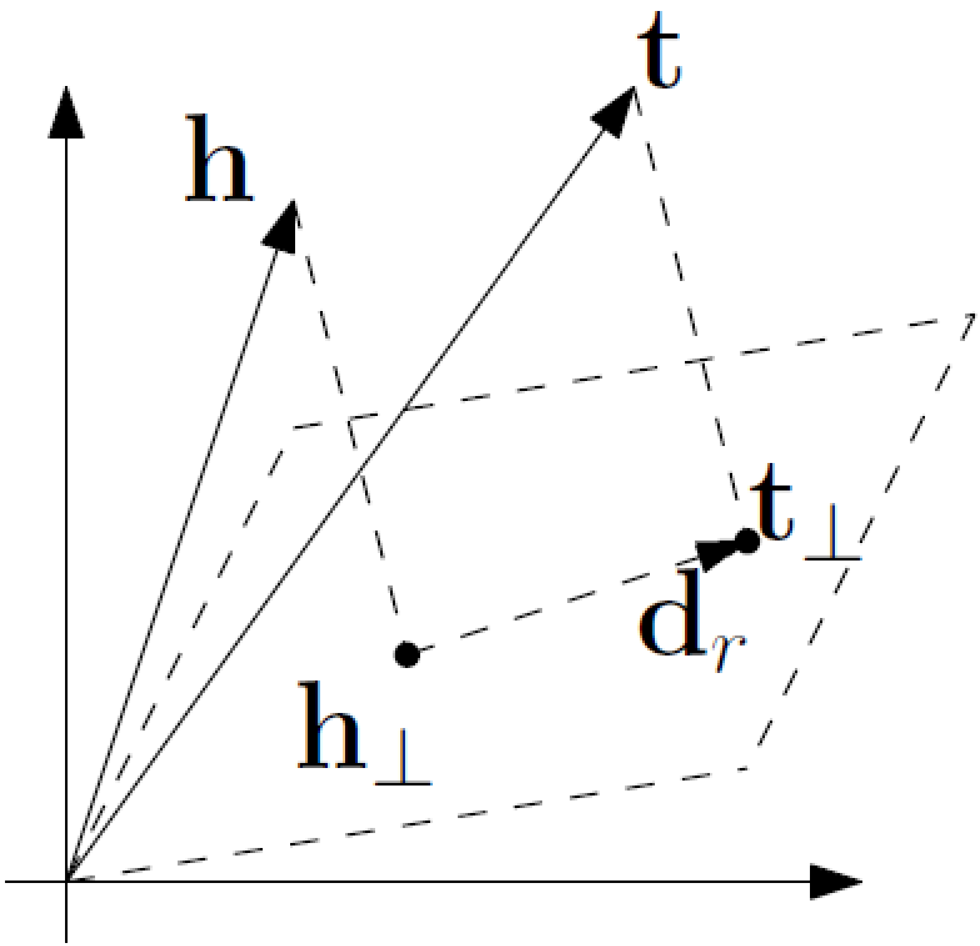
◆score function:

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

◆1-to-1 √

1-to-N/N-to-1/N-to-N ×

TransH

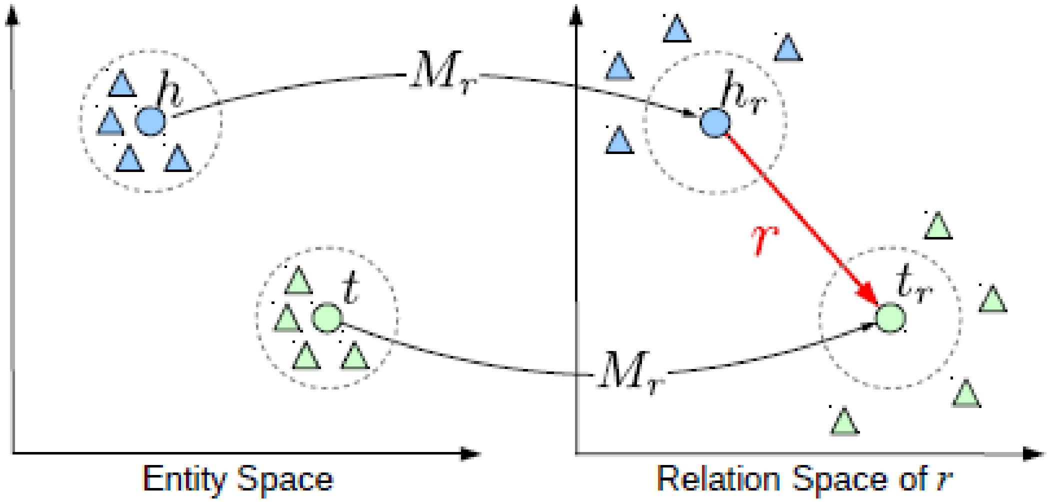


- A triple （ h,r,t ） ,a hyperplane with normal vector ,

- Project h and t to the hyperplane: ,

- Score function:

TransR



- $h, t \in R^k \quad \mathrm{r} \in R^d \quad k \neq d$
- $M_r \in R^{k \times d}$
- Project entities from entity space to relation space:
$$h_r = hM_r \qquad t_r = tM_r$$
- Score function:
$$f_r(h, t) = \|h_r + r - t_r\|_2^2$$
- $\|h\|_2 \leq 1, \|t\|_2 \leq 1, \|r\|_2 \leq 1,$
$\|hM_r\|_2 \leq 1, \|tM_r\|_2 \leq 1$

CTransR: Cluster-based TransR

- TransE/TransH/TransR learn a unique vector for each relation, which may be under-representative to fit all entity pairs under this relation, because these relations are usually rather diverse.

- For a relation r, all entity pairs (h,t) in the training data are clustered into multiple groups, entity pairs in each group are expected to exhibit similar r relation.

- Entity pairs are represented with their vector offsets (h-t), h and t are obtained with TransE.

CTransR: Cluster-based TransR

- For each relation, we learn a separate relation vector r_c for each cluster and M_r.

- $h_{r,c} = hM_r \qquad t_{r,c} = tM_r$

- Score function:

$$f_r(h,t) = \left\| h_{r,c} + r_c - t_{r,c} \right\|_2^2 + \alpha \| r_c - r \|_2^2$$

差别不能太大

Training Objective

$$L = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \max\left(0, f_r(h,t) + \gamma - f_r(h',t')\right)$$

γ is the margin
S is the set f correct triples and
S' is the set of incorrect triples

Data Sets

- WordNet and Freebase
- Freebase
  - （ Steve jobs. Founded, Apple Inc. ）
  - FB15K, FB13
- WordNet
  - Semantic knowledge of words
  - Relations between synsets
  - WN11, WN18

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

Experiment Settings

- Link Prediction
- Triple Classication
- Relation Extraction from Text

Link Prediction

- ◆For each triple (h,r,t), replace the head/tail entity by all entities in the KG.
- ◆Rank these entities in descending (ascending?) order of scores
- ◆Evaluation Metric:

  (1) mean rank of correct entities

  (2) proportion of correct entities in top-10 ranked entities(Hits@10)
- ◆A good link predictor should achieve lower mean rank or higher Hits@10

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**

Triple Classification

- A binary classification: judge whether a given triple (h, r, t) is correct or not.

- A relation-specific threshold δ_r. For a triple (h,r,t), is f_r is below δ_r, the triple will be classified as positive.

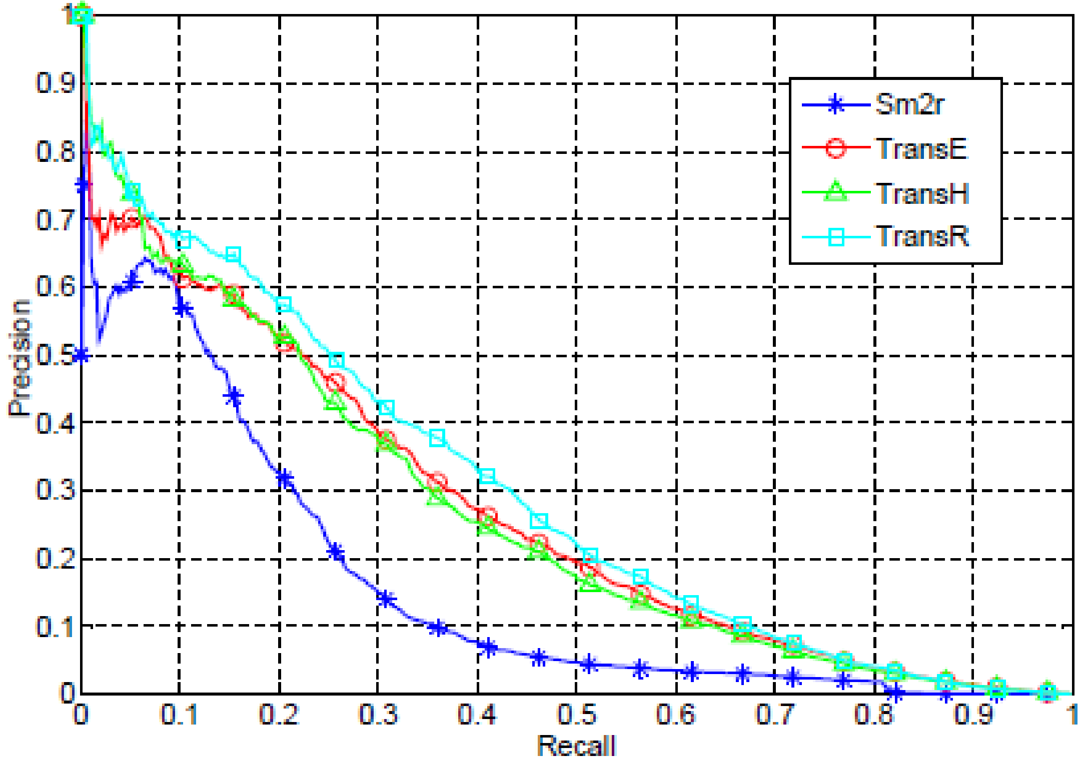
- δ_r is optimized by maximizing classification accuracies on the validation set.

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**	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
TransR (bern)	**85.9**	82.5	83.9
CTransR (bern)	85.7	-	**84.5**

Relation Extraction from Text

- Combine the scores from text-based relation extraction model with the scores from knowledge graph embeddings to rank test triples.

- Plot a precision-recall curves.



Deep Walk

DeepWalk: Online Learning of Social Representations

Bryan Perozzi
Stony Brook University
Department of Computer Science

Rami Al-Rfou
Stony Brook University
Department of Computer Science

Steven Skiena
Stony Brook University
Department of Computer Science

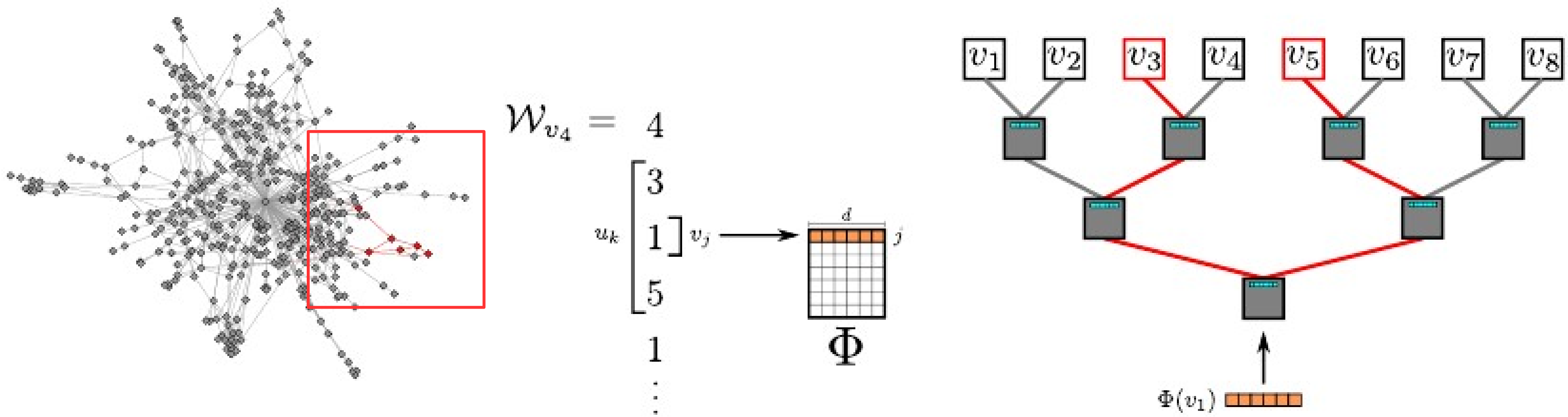
{bperozzi, ralrfou, skiena}@cs.stonybrook.edu

KDD'14

- A graph (e.g., social networks) defines vertexes (analogous to word in word2vec) and edges (analogous to word co-occurrences)
- The basic idea: Train a "LM" over graph nodes

Step I: A small random walk generates context as a "sentence."

Note: The random walk is not actually necessary because we can alternatively define context as adjacent nodes. But as some critics point out, random walk increases randomness and is tunable by its weights.



(a) Random walk generation. (b) Representation mapping. (c) Hierarchical Softmax.

Step II: Standard word2vec (either hierarchical softmax or SkipGram)

An Interesting Phenomenon

- The connections of nodes in social networks, say, typically follow a power law distribution. So does word co-occurrences.