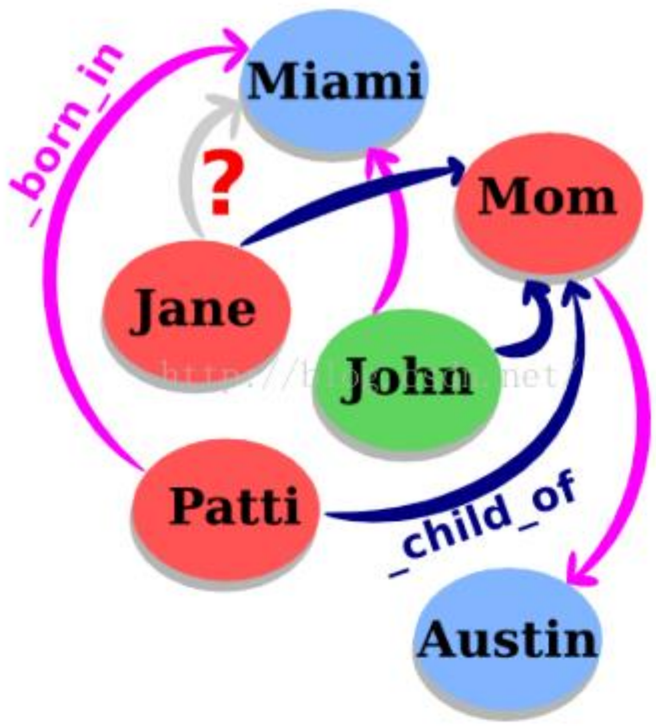
Representation Learning of Knowledge Graphs with Entity Descriptions

Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, Maosong Sun

AAAI-16

Knowledge Graph



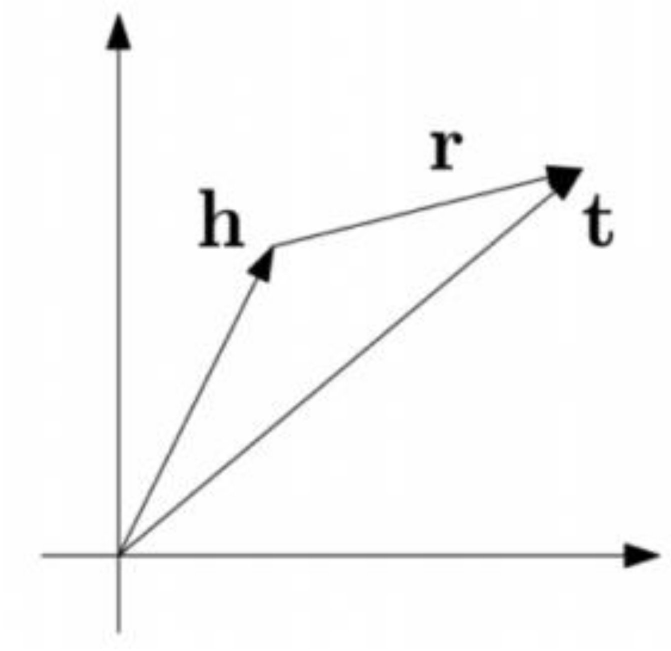
- Triple facts:
- (head entity, relation, tail entity)
- E.g.
- (Jiang Zemin, changed, China)

- Freebase
- Google Knowledge Graph
- DBPedia

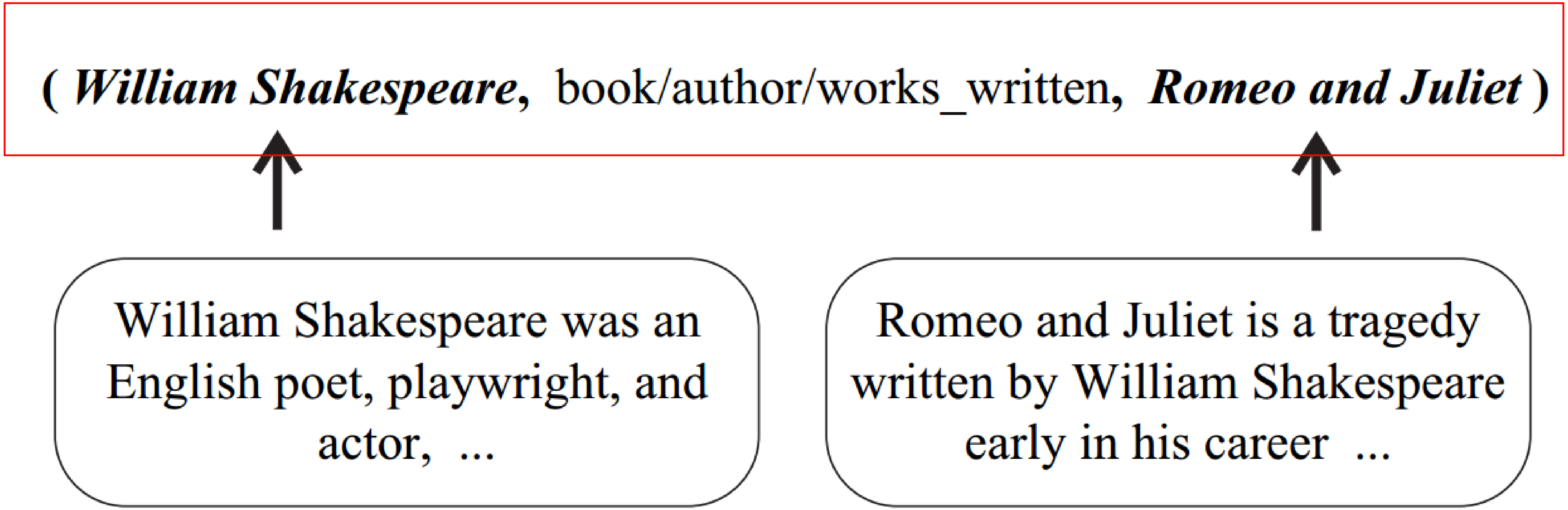
KG Embedding

- As KG size increases, representation learning for KGs has been proposed.
- Entities & relations to low-dimensional vectors
- TransE: relation are translations between head & tail

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



Descriptions

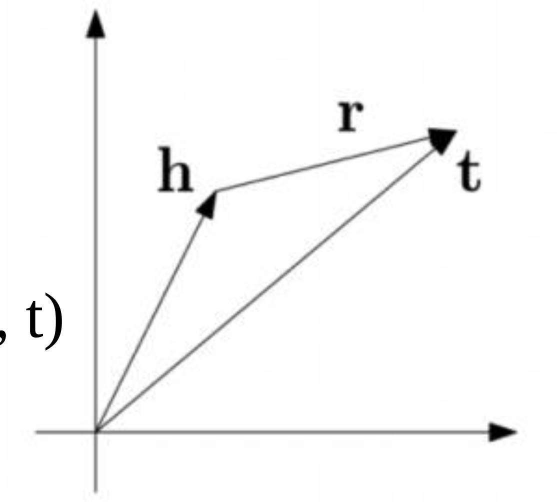


KG Embedding

Structure-based

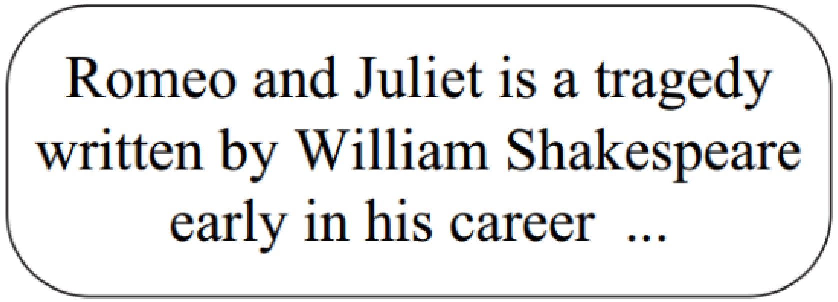
- Capture information in fact triples of KGs

(h, r, t)



Description-based

- Capture textual information in entity descriptions



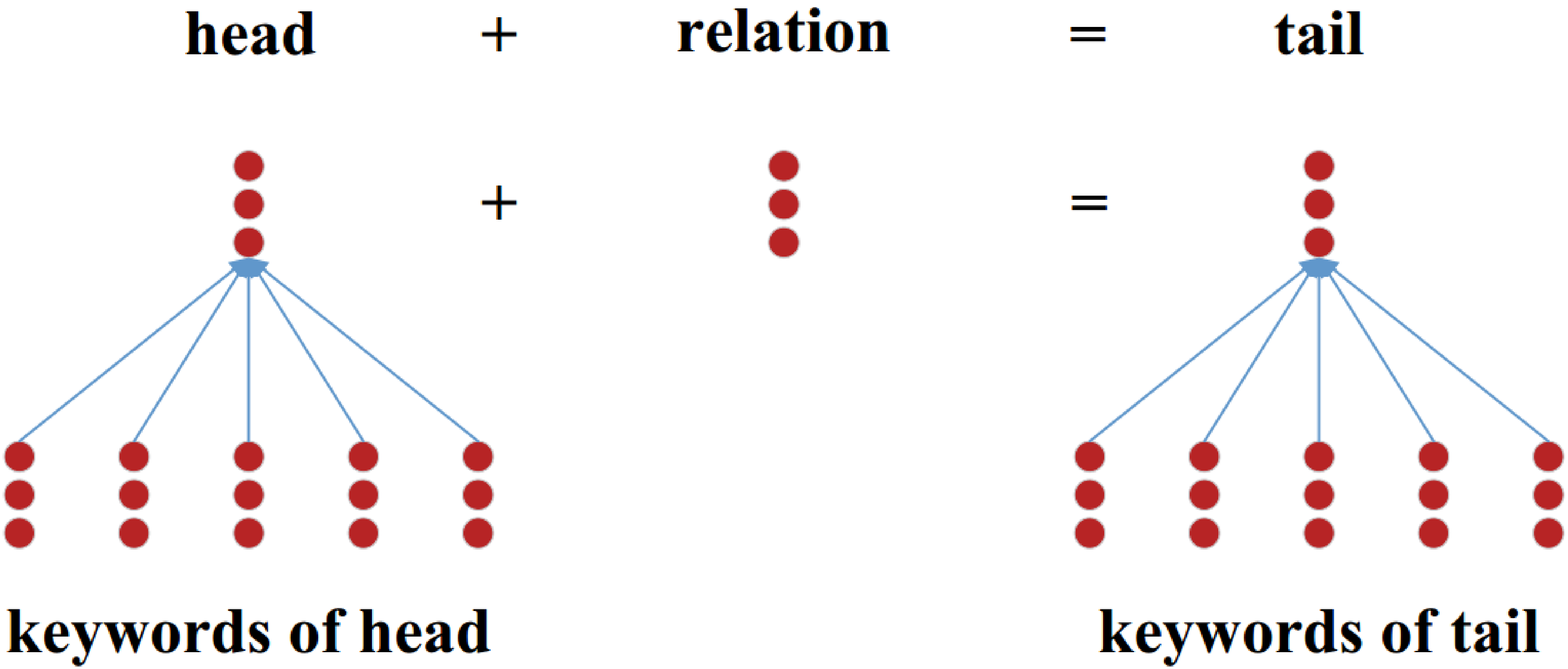
Zero-shot

- When an entity does not occur in training set, how to represent it?
- No structure information could be used
- Existing structure-based methods
- The description of entity should be in consideration

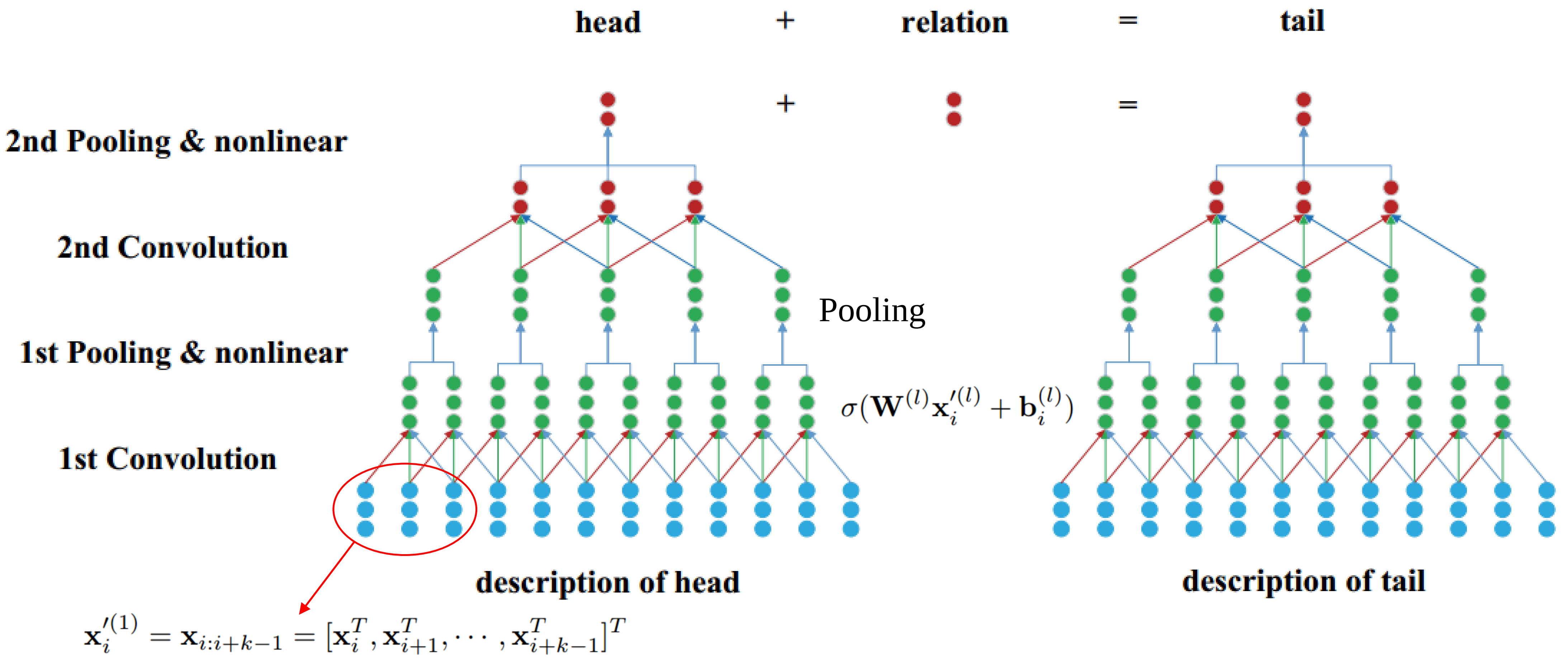
DKRL Model

- Description-Embodied Knowledge Representation Learning
- Take advantages of fact triples and entity description
- Can deal with zero-shot scenario

CBOW



CNN


head
+
relation
=
tail
+
=
2nd Pooling & nonlinear
2nd Convolution
Pooling
1st Pooling & nonlinear
$\sigma(\mathbf{W}^{(l)}\mathbf{x}_i'^{(l)} + \mathbf{b}_i^{(l)})$
1st Convolution
description of head
description of tail
$\mathbf{x}_i'^{(1)} = \mathbf{x}_{i:i+k-1} = [\mathbf{x}_i^T, \mathbf{x}_{i+1}^T, \cdots, \mathbf{x}_{i+k-1}^T]^T$

Pooling

- 1st layer: n-max pooling
- split the output vectors of the convolution layer with size n non-overlapped windows
- Max pooling for each window
- 2nd layer:
- Mean pooling to avoid to much information loss

- Description-based representation is for entities.
- The representations of relations are learned by structure-based methods like TransE.
- Need to map 2 representations to a same space

Energy Functions

- Energy function:

$$E = E_S + E_D$$

$$E_D = E_{DD} + E_{DS} + E_{SD}$$

$$E_S = \|h_s + r - t_s\|$$
$$E_{DD} = \|h_d + r - t_d\|$$
$$E_{DS} = \|h_d + r - t_s\|$$
$$E_{SD} = \|h_s + r - t_d\|$$

Training

- Loss function:

$$L = \sum_{(h,r,t)\in T} \sum_{(h',r',t')\in T'} \max(\gamma + d(h+r,t) - d(h'+r',t'), 0),$$

$$T' = \{(h',r,t)|h' \in E\} \cup \{(h,r,t')|t' \in E\} \cup \{(h,r',t)|r' \in R\},$$

Negative sampling by replace an element of triple

4 Energy functions

$$E_S = \|h_s + r - t_s\|$$
$$E_{DD} = \|h_d + r - t_d\|$$
$$E_{DS} = \|h_d + r - t_s\|$$
$$E_{SD} = \|h_s + r - t_d\|$$

Initialization & Optimization

- X: pre-trained by Word2Vec on Wikipedia
- E, R: initialized randomly or trained by TransE
- W1,W2: initialized randomly

- Back-propagation + Stochastic Gradient Descent

Experiments

- Task:
  - Knowledge graph completion
  - Entity Classification
- Dataset: FB15K&FB20K
- FB15K: an extraction of Freebase
- FB20K: a dataset containing out-of-KG entities created by author

Dataset

Table 1: Statistics of data sets

Dataset	#Rel	#Ent	#Train	#Valid	#Test
FB15K	1,341	14,904	472,860	48,991	57,803

Dataset	#Ent	$\#e-e$	$\#d-e$	$\#e-d$	$\#d-d$
FB20K	19,923	57,803	18,753	11,586	151

Parameters

- entity/relation dimension: {50,80,100}
- learning rate :{0.0005,0.001,0.002}
- margin γ : {0.5,1.0,1.5,2.0}
- CBOW : top-20 keywords
- CNN:
- dimension of word embedding: {50,80,100}
- dimension of feature map: {50,100,150}
- window size k :{1,2,3}

KG Completion

- Complete a triple (h, r, t) when one of h, r, t is missing
- Score function: $\|h + r - t\|$
- Evaluation metrics
- Mean Rank: mean rank of correct entities
- Hits@10(1): proportion of valid entities ranked in top 10. For relation is Hits@1.

KG Completion

Table 2: Evaluation results on entity prediction

Metric	Mean Rank		Hits@10(%)	
	Raw	Filter	Raw	Filter
TransE	210	119	48.5	66.1
DKRL(CBOW)	236	151	38.3	51.8
DKRL(CNN)	200	113	44.3	57.6
DKRL(CNN)+TransE	**181**	**91**	**49.6**	**67.4**

Table 3: Evaluation results on relation prediction

Metric	Mean Rank		Hits@1(%)	
	Raw	Filter	Raw	Filter
TransE	2.91	2.53	69.5	90.2
DKRL(CBOW)	2.85	2.51	65.3	82.7
DKRL(CNN)	2.91	2.55	**69.8**	89.0
DKRL(CNN)+TransE	**2.41**	**2.03**	**69.8**	**90.8**

Entity Classification

- Almost every entity has types in Freebase
- entities in FB15K have 4054 types in Freebase
- select top-50 types with highest frequency : 13445 entities, (12113 for training, 1332 for testing)
- Logistic Regression & one-versus-rest
- Mean Average Precision as evaluation

Entity Classification

Table 4: Evaluation results on entity classification

Metric	FB15K	FB20K
TransE	87.9	-
BOW	86.3	57.5
DKRL(CBOW)	89.3	52.0
DKRL(CNN)	**90.1**	**61.9**

Zero-shot Scenario

- 50 types in FB15K
- 13445 entities in FB15K for training
- 4050 out-of-KG entities for testing

Zero-shot Scenario

Table 5: Evaluation results on entity prediction in zero-shot scenario

Metric	$d-e$	$e-d$	$d-d$	Total
Partial-CBOW	26.5	20.9	67.2	24.6
CBOW	27.1	21.7	66.6	25.3
Partial-CNN	26.8	20.8	69.5	24.8
CNN	**31.2**	**26.1**	**72.5**	**29.5**

Table 6: Evaluation results on relation prediction in zero-shot scenario

Metric	$d-e$	$e-d$	$d-d$	Total
Partial-CBOW	49.0	42.2	0.0	46.2
CBOW	52.2	47.9	0.0	50.3
Partial-CNN	56.6	52.4	4.0	54.8
CNN	**60.4**	**55.5**	**7.3**	**58.2**

Zero-shot Scenario

Table 4: Evaluation results on entity classification

Metric	FB15K	FB20K
TransE	87.9	-
BOW	86.3	57.5
DKRL(CBOW)	89.3	52.0
DKRL(CNN)	**90.1**	**61.9**

Conclusion

- Building representations from entity descriptions achieves better performance than baselines
- Additionally it can deal with out-of-KG entities
- More information of entities could be in consideration
- DKRL can be extended to understand KG structure better by involving extension models like TransH, TransR and PTransE.