LENA: Locality-Expanded Neural Embedding for Knowledge Base Completion

Fanshuang Kong[1], Richong Zhang[1,*], Yongyi Mao[2], Ting Deng[1]

{kongfs, zhangrc, dengting}@act.buaa.edu.cn, ymao@uottawa.ca

[1]School of Computer Science and Engineering, Beihang University, Beijing, China

[2]School of Electrical Engineering & Computer Science, University of Ottawa, Ottawa

Abstract

- Propose a new embedding model and the model assumes that whether a triple is factual depends not only on the embedding of the triple but also on the embeddings of the entities and relations in a larger graph neighborhood.
- Attention mechanisms are constructed to select the relevant information in the graph neighborhood so that irrelevant signals in the neighborhood are suppressed.

Motivation

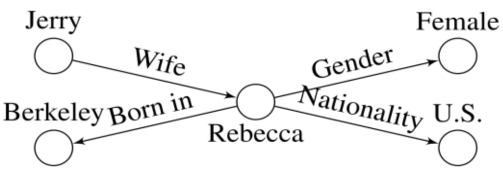

Figure 1: Subgraph of Rebecca

- "Rebecca is the wife of Jerry" is relevant to "Rebecca's gender is female"
- "Rebecca is the wife of Jerry" is not useful for "The nationality of Rebecca is U.S."
- "Rebecca was born in Berkeley" is useful for predicting "The Nationality of Rebecca is U.S."
- "Rebecca was born in Berkeley" is irrelevant to "Rebecca's gender is female"

Insights

- Relationships in the knowledge graph is directed and the direction is meaningful.
- Not all information in a given neighborhood of a triple is relevant to the existence of the triple

An **attention** mechanism, is built in LENA to **(soft-)select** the relevant information in a designated graph neighborhood and suppress the irrelevant noise.

Model

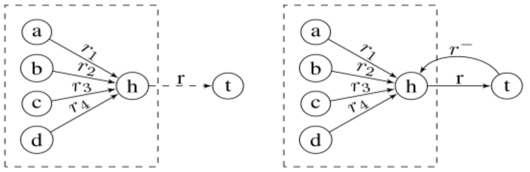

Figure 2: Example of neighborhood graphs $\mathcal{G}(h,r,t)$ (the subgraphs in the dashed boxes) of triple (h,r,t). Triples in $\mathcal{G}$ are represented by a solid edge, and triples(e.g., candidate triples) not in $\mathcal{G}$ are represented by a dashed edge

Window Attention

$$\alpha_e(e_l) := \frac{\exp\langle \gamma_e, r_l\rangle}{\sum_{j=0}^{L}\exp\langle \gamma_e, r_j\rangle}$$

$$\alpha_r(r_l) := \frac{\exp\langle \gamma_r, r_l\rangle}{\sum_{j=0}^{L}\exp\langle \gamma_r, r_j\rangle}$$

Both attention parameters γ_e and γ_r are dependent of the r

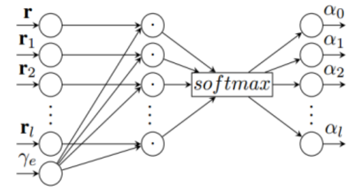

Figure 3: Attention for α_e

Soft-selection

$$v^{\mathrm{E}} := \alpha_e(0)h + \sum_{l=1}^{L}\alpha_e(l)e_l$$

$$v^{\mathrm{R}} := \alpha_r(0)r + \sum_{l=1}^{L}\alpha_r(l)r_l$$

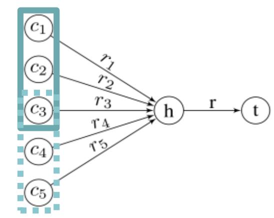

Figure 4: Sample Window

Cross Window Pooling

$$v^{\mathrm{E}}(h,r,t) := \text{max\_pooling}\{v^{\mathrm{E}}(\Gamma) : \Gamma \in \widetilde{\mathcal{H}_L}(h,r,t)\}$$

$$v^{\mathrm{R}}(h,r,t) := \text{max\_pooling}\{v^{R}(\Gamma) : \Gamma \in \widetilde{\mathcal{H}_L}(h,r,t)\}$$

Generate $v^{\mathrm{E}}(h,r,t)$ and $v^{\mathrm{R}}(h,r,t)$ by pooling across the outputs from the windows.

The number H of windows sampled in a neighborhood graph and the window size L are hyperparameters of the model.

Objective function

$$\Theta^* := \arg\min_{\Theta} \sum_{(h,r)\in\mathcal{K}} \sum_{t\in\mathcal{T}(h,r)} \left(-\frac{1}{|\boldsymbol{T}(h,r)|}\log p\,(t|h,r)\right)$$

Experiments / Link Prediction

Table 1: Link Prediction Performance. Superscripts point to the source of reported results.

Models	FB15K-237						WN18-RR					
	MR	FMR	MRR	FMRR	HIT	FHIT	MR	FMR	MRR	FMRR	HIT	FHIT
TransE	367	194	12.1	20.8	28.4	42.0	3542	3529	10.8	12.4	32.9	35.3
TransH	357	186	12.5	21.5	29.3	43.3	3894	3881	11.0	12.7	33.2	35.2
DistMult	453	255/254*	14.0	22.7/24.1*	27.6	40.7/41.9*	7753	7643/5110*	28.1	39.1/43.0*	40.4	41.9/49.0*
ComplEx	456	245/339*	12.8	22.5/24.7*	26.4	41.2/42.8*	8303	8299/5261*	28.1	39.0/44.0*	40.1	41.3/51.0*
Analogy	468	274	14.3	23.3	27.4	40.2	8221	8075	27.6	38.9	39.5	41.0
ProjE	360	193	16.0	29.8	29.3	47.7	3732	3718	27.8	38.2	46.9	50.0
ConvE	483	269/246*	15.3	31.1/31.6*	28.4	48.1/49.1*	4810	4795/5277*	31.1	42.5/46.0*	47.1	49.8/48.0*
R-GCN+	-	-	15.6⋆	24.9⋆	-	41.7⋆	-	-	-	-	-	-
LENA$^{\delta=0.1}$	**328**	**174**	**17.5**	31.0	**32.5**	**49.9**	**3028**	**3014**	28.7	35.7	**48.6**	**51.1**
LENA$^{\delta=0.25}$	**345**	**170**	**16.8**	**31.8**	**31.6**	**50.4**	**3276**	**3262**	30.2	41.5	**48.3**	**51.5**
LENA$^{\delta=0.5}$	364	**175**	**16.3**	**32.0**	**30.8**	**50.4**	**3300**	**3285**	28.3	42.5	**48.5**	**51.4**
Models	**FB15K**						**WN18**					
	MR	FMR	MRR	FMRR	HIT	FHIT	MR	FMR	MRR	FMRR	HIT	FHIT
TransE	194	54	16.6	31.6	48.4	73.9	320	307	28.7	39.3	77.5	92.3
TransH	193	54	16.7	31.9	48.5	74.0	327	314	29.0	39.4	77.8	92.6
DistMult	282	113/97*	24.7/24.2◊	70.8/65.4*	48.9	83.0/82.4*	654	642/902*	52.7/53.2◊	73.9/82.2*	77.6	93.6/93.6*
ComplEx	278	119	25.4/24.2◊	71.6/69.2*	49.9	83.5/84.0*	737	735	64.5/58.7◊	94.2/94.1*	82.2	94.5/94.7*
Analogy	273	114	25.5/25.3+	72.3/72.5*	50.1	83.9/85.4*	725	717	65.6/65.7+	94.2/94.2*	83.3	94.6/94.7*
ProjE	164	53	29.0	62.0	53.8	80.0	281	266	58.1	82.6	81.5	95.2
ConvE	189	48/64*	27.3	69.0/74.5*	52.4	85.4/87.3*	434	417/504*	53.3	94.4/94.2*	79.6	95.5/95.5*
Gaifman	-	75△	-	-	-	84.2△	-	352△	-	-	-	93.9△
R-GCN+	-	-	26.2⋆	69.6⋆	-	84.2⋆	-	-	56.1⋆	81.9⋆	-	96.4⋆
LENA$^{\delta=0.1}$	**153**	50	**30.7**	59.5	**55.9**	79.6	**254**	**242**	**66.4**	89.8	**84.2**	95.6
LENA$^{\delta=0.25}$	**154**	**42**	**29.7**	63.7	**54.7**	81.9	**276**	**261**	65.1	92.7	82.4	95.6
LENA$^{\delta=0.5}$	**161**	**39**	28.6	65.8	53.4	83.1	312	296	62.2	93.8	81.4	95.5

The values without any notation is from our reproduction, the values printed with a single underline are the current "state of the art". The values printed in bold font are results of LENA outperforming this "state of the art". Among them, the top performances are printed in bold font with a double underlines. Source code in https://github.com/fskong/LENA.

Conclusion

- The embeddings of a triple may be insufficient for predicting its factual existence.
- Extracting and combining information from larger graph neighborhoods can therefore improve link-prediction performance.
- We show that attention mechanisms are an effective means of achieving such information extraction and combining.
- LENA has broken a number of performance records, over a range of datasets.

Experiments / Attention

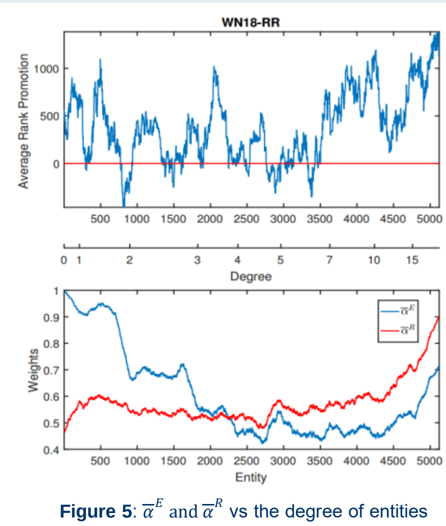

Figure 5: $\overline{\alpha}^E$ and $\overline{\alpha}^R$ vs the degree of entities

Rank Promotion

$$\mathrm{rp}(h,r,t) := rank^{\mathrm{ProjE}}(h,r,t) - rank^{\mathrm{LENA}}(h,r,t)$$

where $rank^{\mathrm{ProjE}}(h,r,t)$ and $rank^{\mathrm{LENA}}(h,r,t)$ are the rank values of (h,r,t) given by ProjE and LENA.

Table 2: Examples of identified informative neighbors

Testing triple	Informative Neighbors	α^E	α^R
Marriott International, Liabilities_Currency, U.S. Dollar	Marriott International, Region, Maryland	0.996	0.501
James Arness, Place_Lived, Minneapolis	James Arness, People_Born_Here, Minneapolis	0.9797	0.0001
Bob Dylan, Instruments_Played, Bass Guitar	Bob Dylan, Instrumentalists, Guitar	0.977	1.59e-06
Hepatitis, Symptom_of, Jaundice	Hepatitis, Risk_Factor, Alcoholism	1.532e-06	0.999