Embedding Models for Knowledge Graphs
Induced by Clusters of Relations and Background Knowledge

RDF Knowledge Graphs & Embedding Models

Knowledge graphs made up of *triples*: $\langle s, p, o \rangle$
    statements with *subject s*, *predicate p*, and *object o*: RDF terms

Embedding Space – given a KG $G = \langle \mathcal{E}_G, \mathcal{R}_G \rangle$ and $k \in \mathbb{N}$ (dim.):
    each *entity* $x \in \mathcal{E}_G$ represented by $\mathbf{e}_x \in \mathbb{R}^k$
    each *predicate* $p \in \mathcal{R}_G$ associated to a **scoring/energy** function

$$f_p : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$$

$f_p(\mathbf{e}_s, \mathbf{e}_o)$ measures the *confidence* in the truth of $\langle s, p, o \rangle$

TransROWL & TransROWL^R [1]

BK Injection

► *triple corruption*: *corrupted* triples Δ' from *observed* triples Δ generated by *reasoning**
  (on domain, range, disjointWith, functionalProperty axioms)

► (explicit) *axiom-based regularization*   *via JENA

Loss functions: to learn the embeddings *margin-based ranking loss* on observed and corrupted triples $\Delta_\pi^{(')}$, $\pi \in \{$inverseOf, equivProperty, equivClass, subClass$\}$

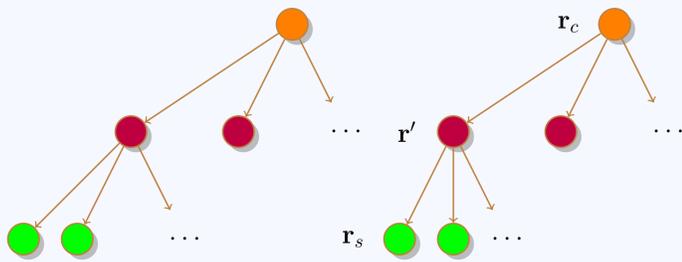see L_B below

Hierarchical Relation Structure [2]

cluster level

relation level

sub-relation level

Loss Function

$$L = L_B + L_{HRS}$$

with: L_B loss function of the *base model*   orig. TRANSE in [2]
and $\mathbf{r} = \mathbf{r}_c + \mathbf{r}' + \mathbf{r}_s$

$$L_{HRS} = \lambda_c \sum_{\mathbf{r}_c \in C} \|\mathbf{r}_c\|_2^2 + \lambda_r \sum_{\mathbf{r}' \in \mathcal{R}} \|\mathbf{r}'\|_2^2 + \lambda_s \sum_{\mathbf{r}_s \in \mathcal{S}} \|\mathbf{r}_s\|_2^2$$

where: C set of clusters or relations, $\mathcal{R} = \mathcal{R}_G$, and $\mathcal{S}$ set of sub-relations

TransROWL-HRS

Top-middle HRS **loss** and **scoring** functions: (top 2 layers in fig.)

$$L_{HRS} = \lambda_c \sum_{\mathbf{r}_c \in C} \|\mathbf{r}_c\|_2^2 + \lambda_r \sum_{\mathbf{r}' \in \mathcal{R}} \|\mathbf{r}'\|_2^2$$

$$f_r(h, t) = \|\mathbf{h}_r + \mathbf{r}_c + \mathbf{r}' - \mathbf{t}_r\|_{1,2}$$

$$f(h, t) = \|\mathbf{h}_r - \mathbf{t}_r\|_{1,2}$$

$\mathbf{h}_r = \mathbf{h}\mathbf{M}_r$ and $\mathbf{t}_r = \mathbf{t}\mathbf{M}_r$

Base **loss** function (via triple *corruption*):   cluster set $C = \{C_1, C_2, \ldots, C_{n_c}\}$

$$L_B = \sum_{c=1}^{n_c} \sum_{r \in C_c} \sum_{\substack{(h,r,t) \in \Delta \\ (h',r',t') \in \Delta'}} [\gamma + f_r(h, t) - f_r(h', t')]_+$$

$$+ \lambda_1 \sum_{c=1}^{n_c} \sum_{q \in C_c} \sum_{\substack{(t,q,h) \in \Delta_{\text{inverseOf}} \\ (t',q,h') \in \Delta'_{\text{inverseOf}}}} [\gamma + f_q(t, h) - f_q(t', h')]_+$$

$$+ \lambda_2 \sum_{c=1}^{n_c} \sum_{s \in C_c} \sum_{\substack{(h,s,t) \in \Delta_{\text{equivProperty}} \\ (h',s,t') \in \Delta'_{\text{equivProperty}}}} [\gamma + f_s(h, t) - f_s(h', t')]_+$$

$$+ \lambda_3 \sum_{c=1}^{n_c} \sum_{\text{typeOf} \in C_c} \sum_{\substack{(h,\text{typeOf},l) \in \Delta_{\text{equivClass}} \\ (h',\text{typeOf},l') \in \Delta'_{\text{equivClass}}}} [\gamma + f_{\text{typeOf}}(h, l) - f_{\text{typeOf}}(h', l')]_+$$

$$+ \lambda_4 \sum_{c=1}^{n_c} \sum_{\text{typeOf} \in C_c} \sum_{\substack{(t,\text{subClassOf},p) \in \Delta_{\text{subClass}} \\ (t',\text{subClassOf},p') \in \Delta'_{\text{subClass}}}} [(\gamma - \beta) + f(t, p) - f(t', p')]_+$$

Algorithm Implementation

Clustering of the relations (top-middle levels of the hierarchy)
► via k-MEANS
Stochastic optimization of the proposed loss function to determine the embeddings
► via SGD | Momentum | AdaGrad

Experiments & Results

4 datasets, 3 flavors, 3 similar models, 2 tasks:

Link Prediction:

	DBpedia15K							
	no typeOf				typeOf			
	MR	H@10	MR	H@10	MR	H@10	MR	H@10
model	(raw)	(raw)	(flt.)	(flt.)	(raw)	(raw)	(flt.)	(flt.)
TRANSR	600.12	60.67	586.83	63.57	504.13	85.01	13.96	95.50
TRANSROWL	606.73	60.59	593.45	63.48	484.04	**85.18**	**13.53**	96.54
TRANSROWL^R	607.43	60.71	594.13	63.65	497.40	85.12	16.50	96.54
TRANSROWL-HRS	600.08	60.62	586.83	63.43	485.08	85.17	14.96	96.61
TRANSROWL-HRS Momentum	605.23	60.48	591.95	63.40	**472.22**	85.12	14.59	**96.85**
TRANSROWL-HRS AdaGrad	**579.47**	**61.18**	**566.21**	**64.00**	506.06	85.01	25.72	94.77

	DBpedia100K							
	no typeOf				typeOf			
	MR	H@10	MR	H@10	MR	H@10	MR	H@10
model	(raw)	(raw)	(flt.)	(flt.)	(raw)	(raw)	(flt.)	(flt.)
TRANSR	2142.10	53.17	2112.42	55.96	**1957.42**	92.04	**1480.26**	92.25
TRANSROWL	2147.56	53.24	2117.87	56.03	1961.75	**92.29**	1503.87	**92.43**
TRANSROWL^R	2121.52	53.08	2091.81	55.95	1971.98	92.24	1511.07	**92.43**
TRANSROWL-HRS	2127.12	52.90	2097.58	55.62	1957.96	92.09	1503.34	92.23
TRANSROWL-HRS Momentum	2126.35	52.88	2096.84	55.58	1970.28	92.16	1526.95	92.31
TRANSROWL-HRS AdaGrad	**2087.74**	**53.47**	**2058.10**	**56.25**	1960.09	92.10	1519.92	92.23

	DBpediaYAGO							
	no typeOf				typeOf			
	MR	H@10	MR	H@10	MR	H@10	MR	H@10
model	(raw)	(raw)	(flt.)	(flt.)	(raw)	(raw)	(flt.)	(flt.)
TRANSR	7271.50	**44.64**	7239.09	**46.07**	844.51	81.98	348.65	88.99
TRANSROWL	7209.02	44.45	7176.64	45.84	868.27	82.81	373.90	91.17
TRANSROWL^R	7226.55	44.13	7194.21	45.52	845.42	81.71	352.16	88.77
TRANSROWL-HRS	7189.25	44.42	7156.92	45.68	705.91	82.96	213.18	90.98
TRANSROWL-HRS Momentum	7144.80	44.26	7112.47	45.59	731.13	82.82	241.69	92.59
TRANSROWL-HRS AdaGrad	**7104.88**	44.61	**7072.72**	45.92	**642.92**	**84.85**	**176.32**	**95.20**

	NELL							
	no typeOf				typeOf			
	MR	H@10	MR	H@10	MR	H@10	MR	H@10
model	(raw)	(raw)	(flt.)	(flt.)	(raw)	(raw)	(flt.)	(flt.)
TRANSR	**6891.20**	**47.40**	**6681.76**	**55.93**	2315.08	79.94	2140.16	80.50
TRANSROWL	7136.77	46.72	6929.10	55.40	2334.50	80.00	2161.67	80.56
TRANSROWL^R	7339.53	46.09	7132.22	54.15	**2310.11**	79.52	**2138.99**	80.21
TRANSROWL-HRS	7203.32	46.83	6996.17	55.34	2397.16	79.95	2223.64	80.44
TRANSROWL-HRS Momentum	7144.77	46.87	6936.06	55.29	2373.70	79.29	2201.06	79.85
TRANSROWL-HRS AdaGrad	7153.78	46.67	6949.22	55.26	2903.37	**80.35**	2730.42	**80.80**

Triple Classification:

	DBpedia15K							
	no typeOf				typeOf			
model	Acc.	P	R	FPR	Acc.	P	R	FPR
TRANSR	**0.641**	0.998	0.364	**0.001**	0.972	0.966	0.946	0.378
TRANSROWL	0.631	0.997	0.347	0.002	0.962	**0.999**	0.882	**0.006**
TRANSROWL^R	0.628	**0.998**	0.342	**0.001**	0.981	0.969	0.972	0.523
TRANSROWL-HRS	0.634	0.995	0.353	0.002	0.985	0.994	0.961	0.135
TRANSROWL-HRS Momentum	0.628	0.997	0.342	**0.001**	**0.988**	0.997	0.966	0.074
TRANSROWL-HRS AdaGrad	0.629	**0.998**	0.343	**0.001**	0.977	0.954	**0.978**	0.682

	DBpedia100K							
	no typeOf				typeOf			
model	Acc.	P	R	FPR	Acc.	P	R	FPR
TRANSR	0.711	0.998	0.313	0.001	0.976	0.884	0.800	**0.344**
TRANSROWL	0.705	**0.998**	0.300	**0.001**	**0.987**	**0.940**	0.895	0.353
TRANSROWL^R	0.704	0.998	0.298	0.001	0.981	0.872	0.890	0.543
TRANSROWL-HRS	**0.714**	**0.998**	**0.320**	**0.001**	0.979	0.900	0.832	0.353
TRANSROWL-HRS Momentum	0.709	0.997	0.310	0.001	0.985	0.874	**0.945**	0.711
TRANSROWL-HRS AdaGrad	0.693	**0.998**	0.270	**0.001**	0.977	0.885	0.816	0.367

	DBpediaYAGO							
	no typeOf				typeOf			
model	Acc.	P	R	FPR	Acc.	P	R	FPR
TRANSR	0.644	0.964	0.300	0.016	0.844	0.946	0.247	0.018
TRANSROWL	**0.649**	0.968	**0.307**	0.014	0.905	0.973	0.547	0.032
TRANSROWL^R	0.636	**0.981**	0.277	**0.007**	0.854	0.953	0.299	0.020
TRANSROWL-HRS	0.643	0.966	0.296	0.015	0.873	0.967	0.386	0.021
TRANSROWL-HRS Momentum	0.645	0.974	0.299	0.011	0.873	0.981	0.381	**0.012**
TRANSROWL-HRS AdaGrad	0.647	0.975	0.302	0.011	**0.950**	**0.986**	**0.765**	0.043

	NELL							
	no typeOf				typeOf			
model	Acc.	P	R	FPR	Acc.	P	R	FPR
TRANSR	**0.758**	0.843	**0.636**	0.245	0.803	0.389	0.519	0.630
TRANSROWL	0.744	0.835	0.608	0.234	0.763	0.334	0.560	0.717
TRANSROWL^R	0.739	0.845	0.587	0.207	0.760	0.337	0.598	0.745
TRANSROWL-HRS	0.735	0.818	0.603	0.252	0.748	0.330	0.633	0.778
TRANSROWL-HRS Momentum	0.741	0.810	0.628	0.284	**0.823**	0.421	0.437	**0.516**
TRANSROWL-HRS AdaGrad	0.740	**0.847**	0.585	**0.202**	0.667	0.269	**0.692**	0.859

Possible Extensions

► better/specific search of the hyperparameter values

► extension of the loss function to the whole hierarchical structure
  (including the middle-bottom level)

► replacement of the core embedding model with a more complex one

References

[1] C. d'Amato, N. Quatraro, and N. Fanizzi. Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In *ESWC 2021*, volume 12731 of *LNCS*, pages 441–457. Springer, 2021.

[2] Z. Zhang, F. Zhuang, M. Qu, F. Lin, and Q. He. Knowledge graph embedding with hierarchical relation structure. In *EMNLP 2018*, pages 3198–3207. ACL, 2018.

Claudia d'Amato, Nicola F. Quatraro, and Nicola Fanizzi

{claudia.damato,nicola.fanizzi}@uniba.it n.quatraro@studenti.uniba.it

ARA/LACAM – Dipartimento di Informatica – CILA – Università degli Studi di Bari Aldo Moro