Holographic Embeddings of Knowledge Graph

Maximilian Nickel, Lorenzo Rosasco and Tomaso Poggio

MIT & Cambridge

April 1, 2016

Related Work

- Based on **Compositional Vector Space**.
- What's **Compositional Vector Space**.

$$
\begin{aligned}
\mathcal{P}((\mathbf{h}, \mathbf{r}, \mathbf{t}) | \Theta) \quad & = \sigma(\eta_{h,r,t}) \\
& = \sigma(\mathbf{r}^\top (\mathbf{h} \circ \mathbf{t}))
\end{aligned} \tag{1}
$$

- The critical operator $\circ$:
  - **Tensor Product**:

    $$\mathbf{a} \circ \mathbf{b} = \mathbf{a} \otimes \mathbf{b}$$

    $$f_r(h, t) = \mathbf{r}^\top (\mathbf{h} \otimes \mathbf{t}) = \mathbf{h}^\top \mathbf{R} \mathbf{t}$$

    But it is difficult to identify the head and tail.
  - **Concatenation, Projection and Non-Linearity**:

    $$\mathbf{a} \circ \mathbf{b} = \psi(\mathbf{W}[\mathbf{a}, \mathbf{b}])$$

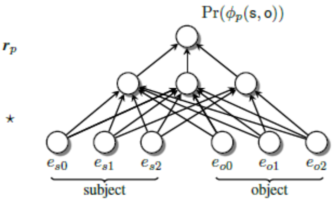
    This branch is sort of over-fitting, and is nearly abandoned.

Holographic Embeddings

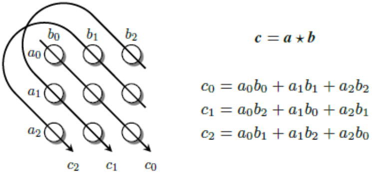
- **Motivation**: To combine the expressive power of the tensor product with the efficiency and simplicity of TransE.
- Compositional Operator (Circular Correlation):

$$\mathbf{a} \circ \mathbf{b} = \mathbf{a} \star \mathbf{b}$$

$$[\mathbf{a} \star \mathbf{b}]_k = \sum_{i=0}^{d-1} a_i b_{(k+i)_{\%d}}$$



(b) HOLE



Holographic Embeddings

- **Associative Memory**: Similar to the associative memory model but instead adopting the correlation operator.

$$\frac{\textit{Paper} : t = \arg\max_{||\mathbf{e_i}||=1} \mathbf{e_i}^\top (\mathbf{r} \star \mathbf{m}_t)}{\textit{Mine} : t = \arg\max_{||\mathbf{e_i}||=1} \mathbf{e_i}^\top \mathbf{m}_t}, \quad \mathbf{m}_t = \sum_{(\mathbf{h},\mathbf{r},t)\in\Delta} \mathbf{r} * \mathbf{h}$$

- Proof:
  1. Regarding the objective as

$$\mathbf{J} = \sum_{(\mathbf{h},\mathbf{r},\mathbf{t})\in\boldsymbol{\Delta}} \mathbf{r}^\top (\mathbf{h} \star \mathbf{t}) - \lambda \sum_{\mathbf{e}\in\mathbf{E}} ||\mathbf{e}||_2^2$$

  2. Thus, let the derivative $\nabla_t \mathbf{J} = \mathbf{0}$, then

$$t = \sum_{(\mathbf{h},\mathbf{r},t)\in\boldsymbol{\Delta}} \mathbf{r} * \mathbf{h} = \mathbf{m}_t$$

  3. Correspondingly, $t = \arg\max \mathbf{e_i}^\top t = \arg\max \mathbf{e_i}^\top \mathbf{m}$.

Experiments

	WN18					FB15k				
	MRR		Hits at			MRR		Hits at		
Method	Filter	Raw	1	3	10	Filter	Raw	1	3	10
TRANSE	0.495	0.351	11.3	88.8	94.3	0.463	0.222	29.7	57.8	**74.9**
TRANSR	0.605	0.427	33.5	87.6	94.0	0.346	0.198	21.8	40.4	58.2
ER-MLP	0.712	0.528	62.6	77.5	86.3	0.288	0.155	17.3	31.7	50.1
RESCAL	0.890	0.603	84.2	90.4	92.8	0.354	0.189	23.5	40.9	58.7
HOLE	**0.938**	**0.616**	**93.0**	**94.5**	**94.9**	**0.524**	**0.232**	**40.2**	**61.3**	73.9

	Countries		
	AUC-PR		
Method	S1	S2	S3
Random	0.323	0.323	0.323
Frequency	0.323	0.323	0.308
ER-MLP	0.960	0.734	0.652
RESCAL	**0.997**	0.745	0.650
HOLE	**0.997**	**0.772**	**0.697**

Approximated Equivalence Between TransE and HOLE

Conclusions: The entity representations between TransE and HOLE are approximately isomorphic, leading to an approximated equivalence.

$$\mathbf{Ent}_{TransE} = \ln \circ \mathbf{Im} \circ \mathcal{F} \circ \mathbf{Ent}_{HOLE}$$

Proof:

$$\begin{aligned}
\mathbf{r}^\top [\mathbf{h} \otimes \mathbf{t}] \quad & \approx \mathcal{F}_{Ent}^{-1} \circ \left\{ \mathbf{r}^\top \left[\hat{\mathbf{h}} \odot \bar{\hat{\mathbf{t}}} \right] \right\} \\
& \overset{\mathbf{Im}(\hat{\mathbf{e}})=\exp(\mathbf{e}')}{\underset{\mathbf{r}=\exp(\mathbf{r}')}{=\!=\!=\!=\!=}} \mathcal{F}_{Ent}^{-1} \circ \left\{ \sum_i exp(\mathbf{h} + \mathbf{r} - \mathbf{t})_\mathbf{i} \right\} \\
& = \mathbf{Re} \left\{ \sum_i exp((\mathbf{h} + \mathbf{r} - \mathbf{t})\mathbf{j})_\mathbf{i} \right\} \\
& = \sum_i \cos(\mathbf{h} + \mathbf{r} - \mathbf{t})_i \\
& \approx -||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2^2 \qquad (2)
\end{aligned}$$

Conclusion

- **HOLE** is basically a translation-based method, approximately equivalent to **TransE**.
- But with a new form, the optimization and parameter-turing are supposed to be different, maybe more effective.
- Besides, this paper casts an associated memory perspective for HOLE(TransE).
- Though, it's a poorly-written paper.

Thanks.

SSP:Semantic Space Projection

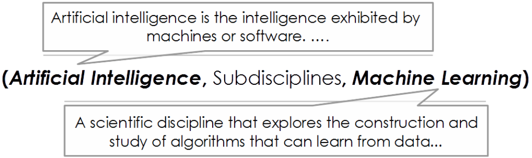
Han Xiao, Minlie Huang and Xiaoyan Zhu

Tsinghua University

April 1, 2016

SSP: Semantic Space Projection

- **Problem Definition** To incorporate the textual descriptions with the fact triples.

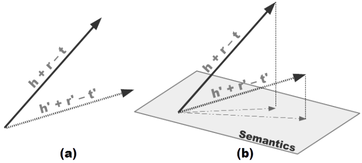


- **Motivations**
  - Discovering semantic relevance between entities.
  - Offering precise semantic expression.
- **Related Work** could not characterize the correlations.
  - Jointly: $\mathbf{w} = \mathbf{e}$.
  - DKRL: $[\mathbf{e_h}, \mathbf{w_h}] \xrightarrow{\mathbf{r}} [\mathbf{e_t}, \mathbf{w_t}]$.

SSP: Semantic Space Projection

- **Methodology**: Projecting the embedding procedure onto a semantic hyperplane.

$$f_r(h, t) = -\lambda||\mathbf{e} - \mathbf{s}^\top \mathbf{e}\mathbf{s}||_2^2 + ||\mathbf{e}||_2^2$$



- **Semantic Vector Generation**: Topic Model.
- **Objectives.**

$$\mathcal{L} = \mathcal{L}_{embed} + \mu\mathcal{L}_{topic}$$

$$\mathcal{L}_{embed} = \sum_{\substack{(h,r,t)\in\Delta \\ (h',r',t')\in\Delta'}} [f_{r'}(h', t') - f_r(h, t) + \gamma]_+$$

$$\mathcal{L}_{topic} = \sum_{e\in E,\ w\in D_e} (C_{e,w} - \mathbf{s}_\mathbf{e}^\top \mathbf{w})^2 \tag{3}$$

SSP: Semantic Space Projection

- **Methodology**: Projecting the embedding procedure onto a semantic hyperplane.

$$f_r(h, t) = -\lambda||\mathbf{e} - \mathbf{s}^\top \mathbf{e}\mathbf{s}||_2^2 + ||\mathbf{e}||_2^2$$

- **Correlation Perspective**: There exists the important restriction, that the entities co-occur in a triple should be embedded in the semantic space composed by the associated textual semantics.
- **Semantic Perspective**: Our model characterizes the strong correlations with a semantic hyperplane, which is capable of taking the advantages of both two semantic effects.
  - Semantic Relevance.
  - Precise Semantic Expression.

SSP: Semantic Space Projection

- **Experiments**: Knowledge Graph Completion.

FB15K	Mean Rank		HITS@10	
TransE	210	119	48.5	66.1
TransH	212	87	45.7	64.4
Jointly	167 [1]	39 [1]	51.7 [1]	77.3 [1]
DKRL(BOW)	200	113	44.3	57.6
DKRL(ALL)	181	91	49.6	67.4
SSP (Std.)	**154**	**77**	57.1	78.6
SSP (Joint)	163	82	**57.2**	**79.0**
WN18	**Mean Rank**		**HITS@10**	
TransE	263	251	75.4	89.2
TransH	401	338	73.0	82.3
SSP (Std.)	204	193	81.3	91.4
SSP (Joint)	**168**	**156**	**81.2**	**93.2**

SSP: Semantic Space Projection

- **Experiments**: Entity Classification.

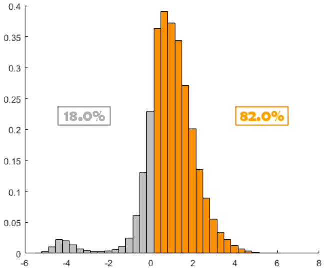
Metrics	FB15K	FB20K
TransE	87.8	-
BOW	86.3	57.5
DKRL(BOW)	89.3	52.0
DKRL(ALL)	90.1	61.9
NMF	86.1	59.6
SSP (Std.)	93.2	-
SSP (Joint)	**94.4**	**67.4**

SSP: Semantic Space Projection

- Semantic Relevance Analysis

	$\mathbf{SSP(S.)}_{\# \leq 100}$	$\mathbf{SSP(J.)}_{\# \leq 100}$
$\mathbf{E}_{\# \geq 500}$	601	672
$\mathbf{E}_{\# \geq 1000}$	275	298
$\mathbf{E}_{\# \geq 2000}$	80	89
$\mathbf{E}_{\# \geq 3000}$	32	39
$\mathbf{E}_{\# \geq 5000}$	3	3

- Precise Semantic Expression Analysis



SSP: Semantic Space Projection

- **Conclusion.**
  - In this paper, we propose the knowledge graph embedding model SSP, which jointly learns from the symbolic triples and textual descriptions.
  - SSP could interact the triples and texts by characterizing the strong correlations, by which means, the textual descriptions could make more effects to discover semantic relevance and offer precise semantic expression.
  - Extensive experiments show our method achieves the substantial improvements against the state-of-the-art baselines.

Thanks.