Joint Language Semantic and Structure Embedding for Knowledge Graph Completion

COLING 2022

Jianhao Shen, Chenguang Wang, Linyuan Gong, Dawn Song

Knowledge Graph

Knowledge Graph

Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)

Knowledge Graph

Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)

Instances

Freebase

WIKIDATA

yago
select knowledge

WORDNET

Knowledge Graph

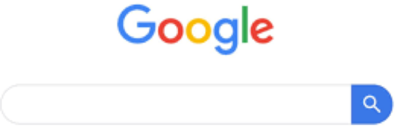
Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)

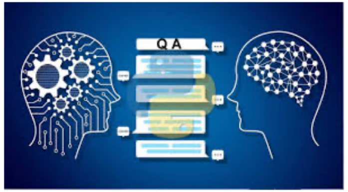
Instances

Freebase WIKIDATA yago select knowledge WORDNET

Applications



Web Search



QA System

Dialogue System

Knowledge Graph

Form

(head entity, relation, tail entity)
(Steve Jobs, founder of, Apple Inc.)

Instances

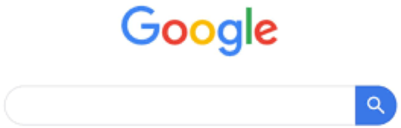
Freebase

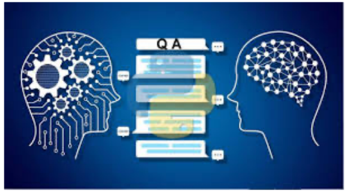
WIKIDATA

yago
select knowledge

WORDNET

Applications





Web Search

QA System

Dialogue System

Knowledge graphs are crucial to many NLP applications

Knowledge Graph Completion

Incompleteness impedes KG’s adoption in real-world applications

Knowledge Graph Completion

Incompleteness impedes KG’s adoption in real-world applications

Knowledge Graph Completion

Incompleteness impedes KG's adoption in real-world applications

Link Prediction: (h, r, ?)

Triplet Classification: (h, r, t) true or false?

Knowledge Graph Completion

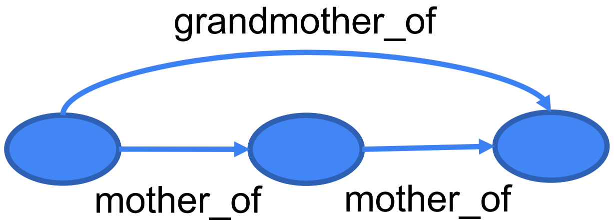
Incompleteness impedes KG’s adoption in real-world applications

Link Prediction: (h, r, ?)

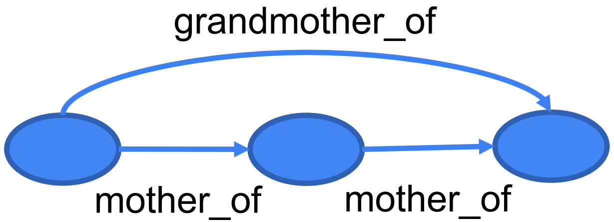
Triplet Classification: (h, r, t) true or false?

How to represent triplets and infer the missing ones?

Structural Pattern



Structural Pattern



Structure-based approaches:

- Treat entities and relations as nodes and edges
- Use graph embedding
- TransE, TransH, RotatE

Semantic Information

Bill Gates started Microsoft in 1975 with Paul Allen

→ (Bill Gates, founder_of, Microsoft)

Semantic Information

Bill Gates started Microsoft in 1975 with Paul Allen

→ (Bill Gates, founder_of, Microsoft)

Semantic-based approaches:

- Leverage text description for entities and relations
- Use language modeling
- KG-BERT

Semantic Information

Bill Gates started Microsoft in 1975 with Paul Allen

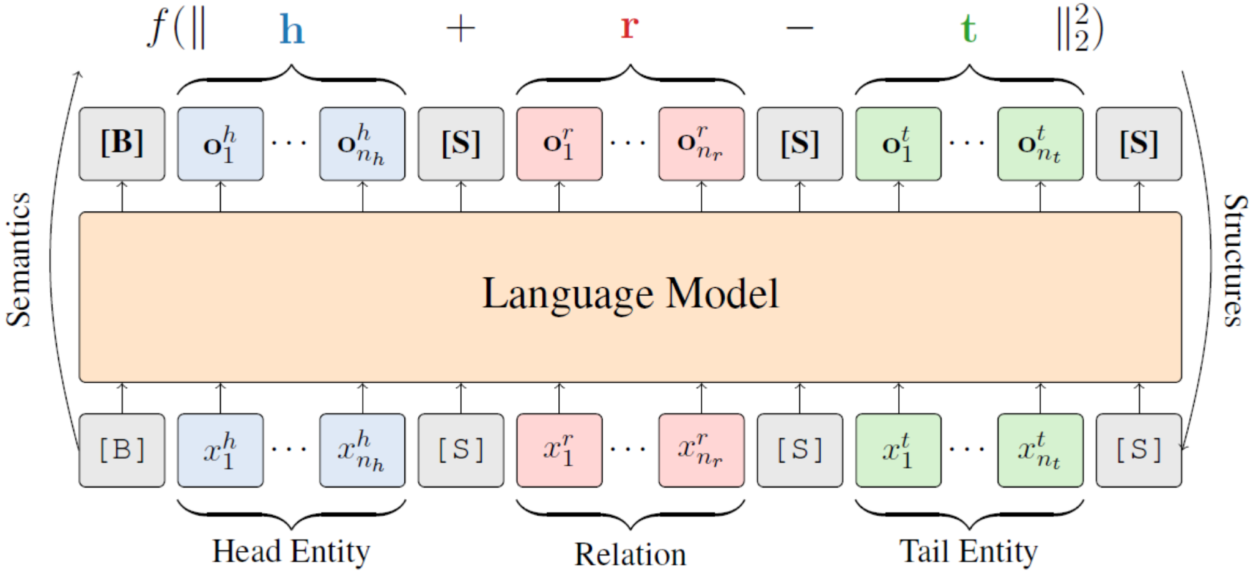
→ (Bill Gates, founder_of, Microsoft)

Semantic-based approaches:

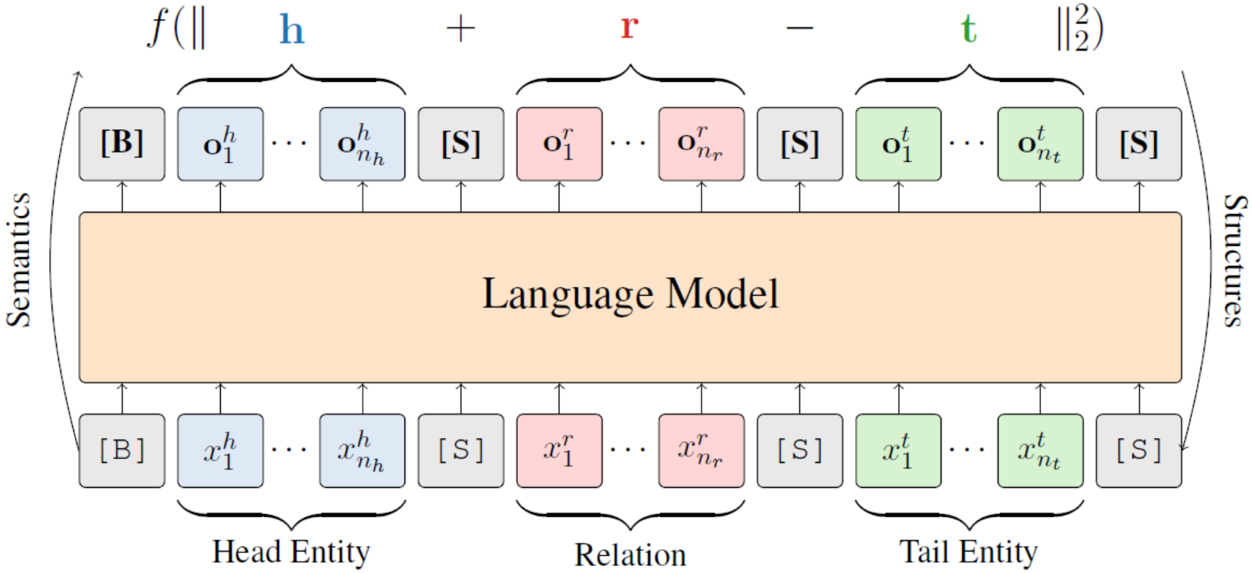
- Leverage text description for entities and relations
- Use language modeling
- KG-BERT

Both structures and semantics are vital to KG completion

Joint **La**nguage **S**emantic and **S**tructure Embedding (LASS)



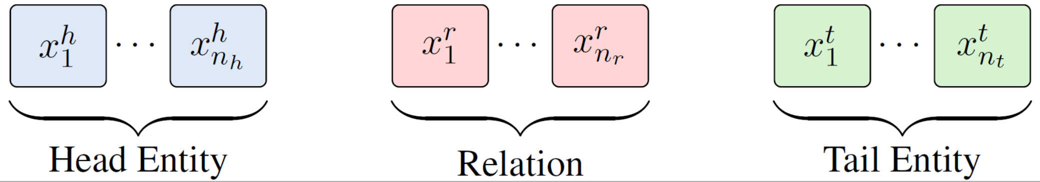
Joint **La**nguage **S**emantic and **S**tructure Embedding (LASS)



Key idea: embeds a triplet by fine-tuning PLMs w.r.t. a structured loss

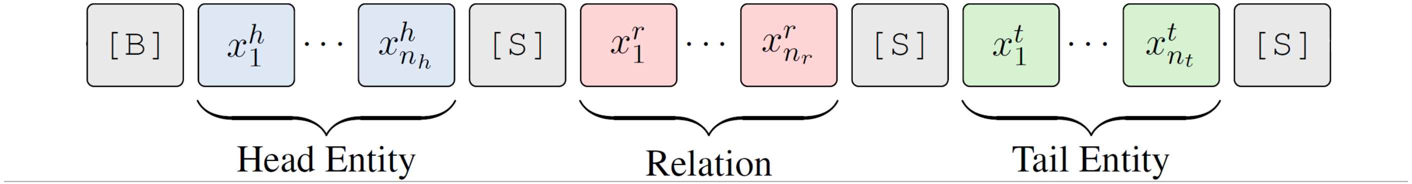
Semantic Embedding

Semantic Embedding

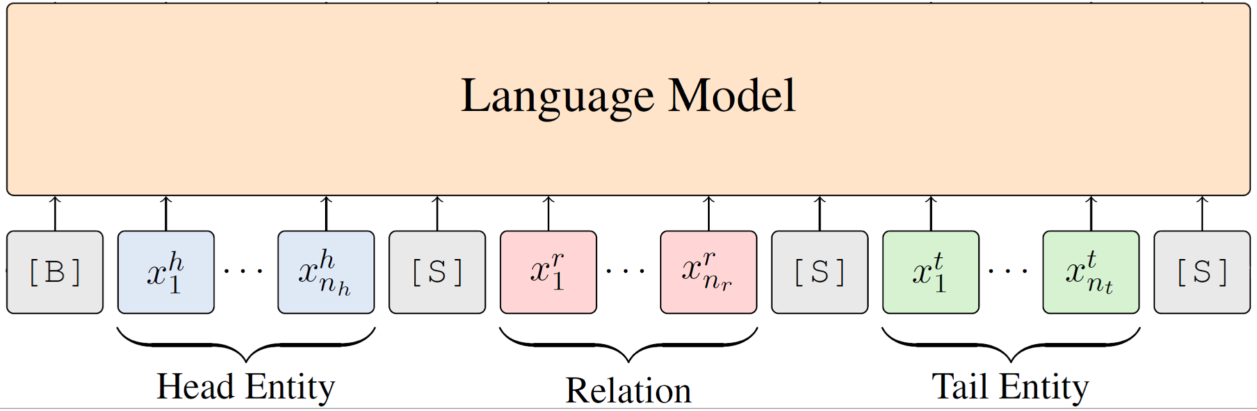


$$\underbrace{x_1^h \cdots x_{n_h}^h}_{\text{Head Entity}} \quad \underbrace{x_1^r \cdots x_{n_r}^r}_{\text{Relation}} \quad \underbrace{x_1^t \cdots x_{n_t}^t}_{\text{Tail Entity}}$$

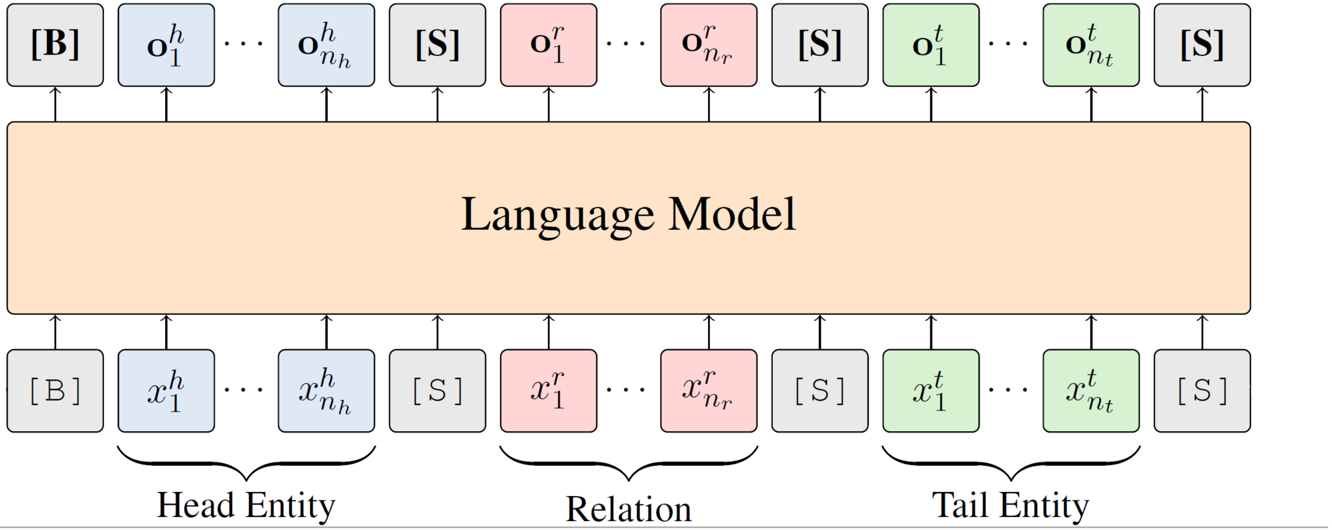
Semantic Embedding



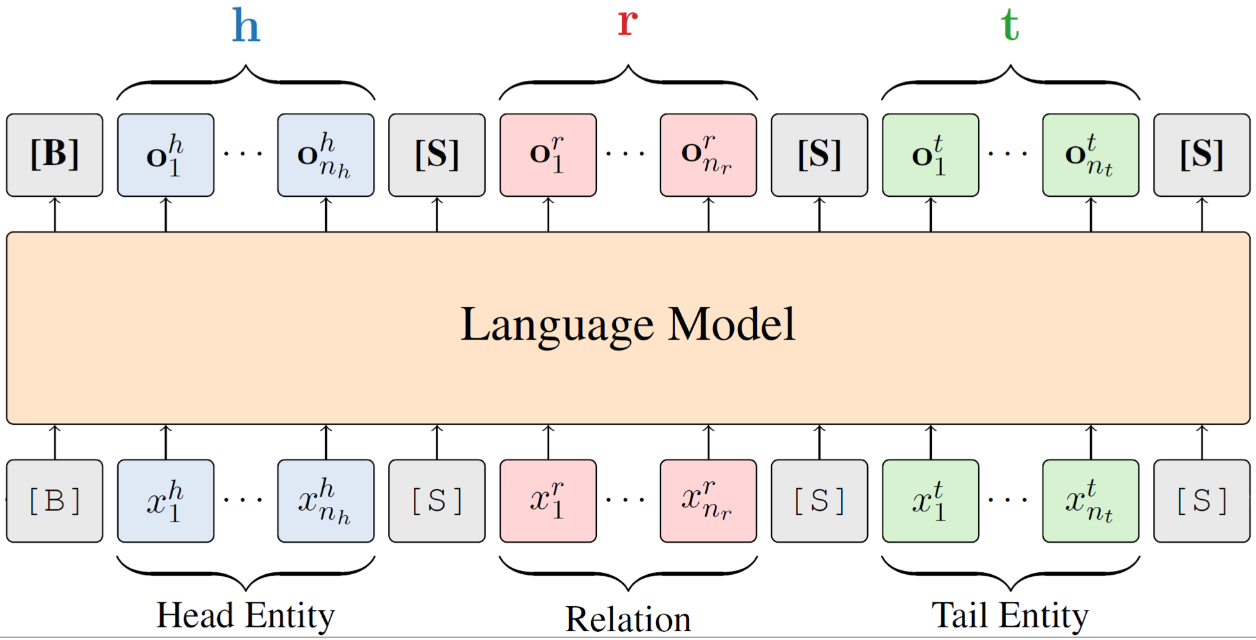
Semantic Embedding



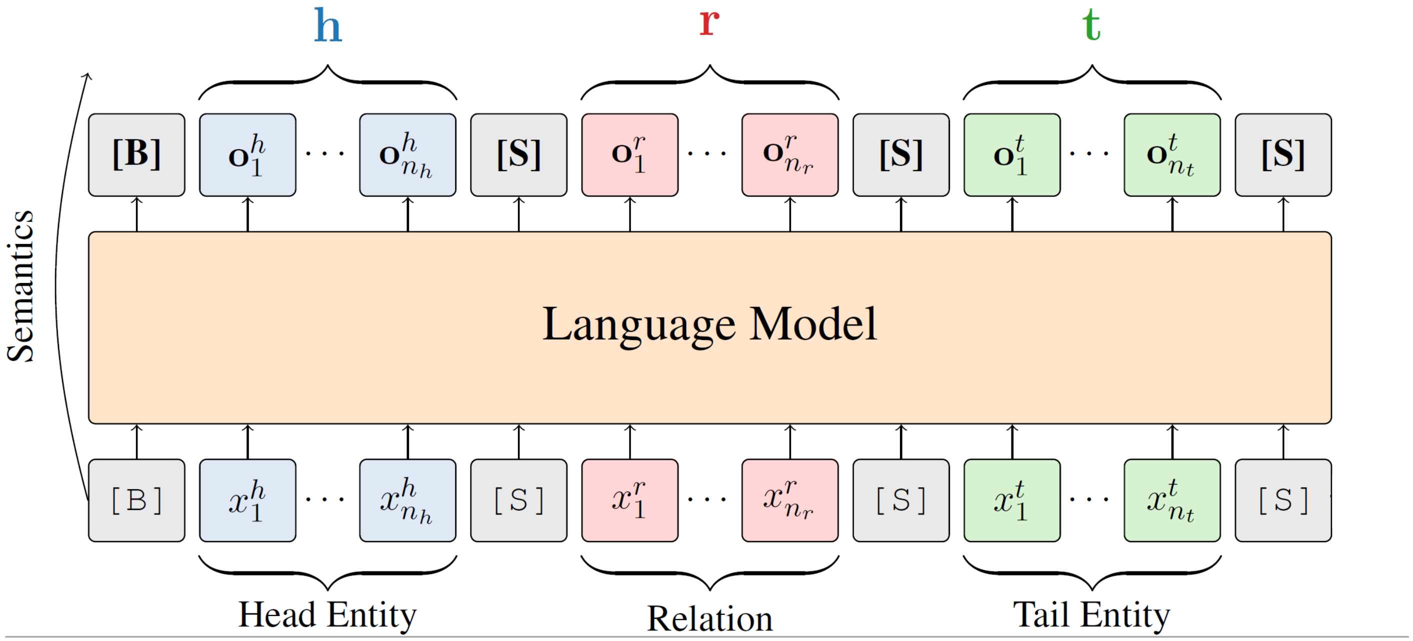
Semantic Embedding



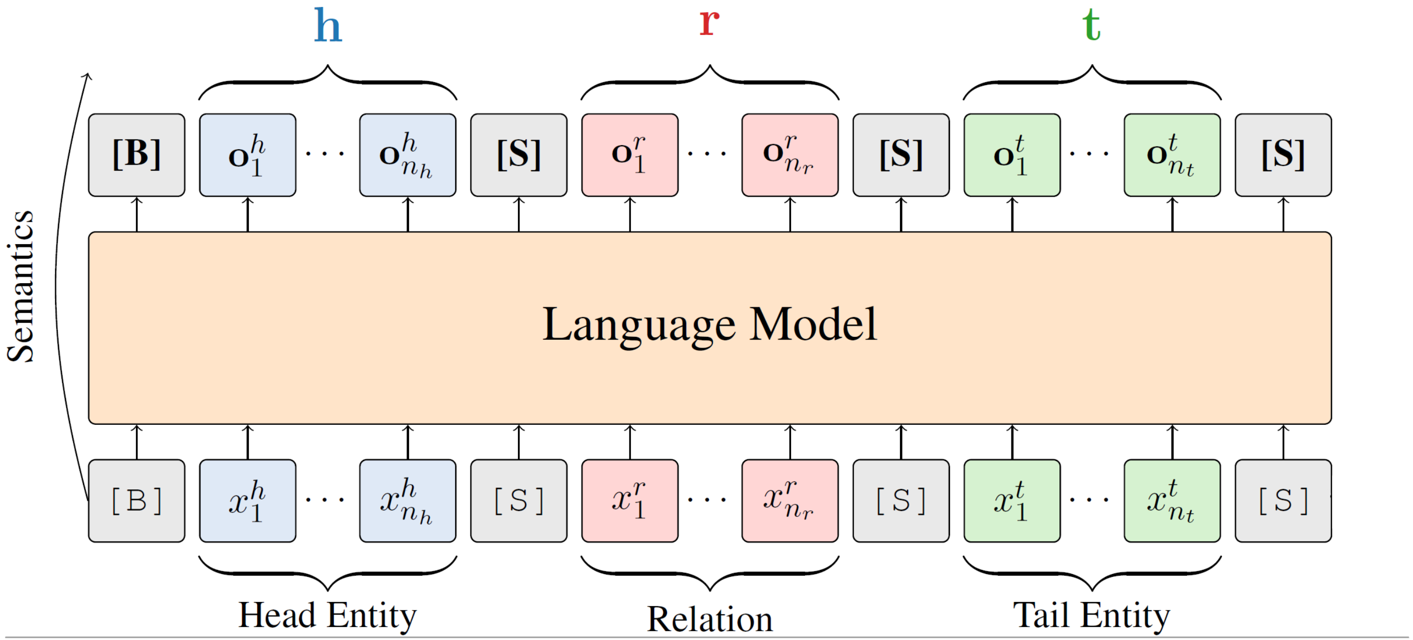
Semantic Embedding



Semantic Embedding

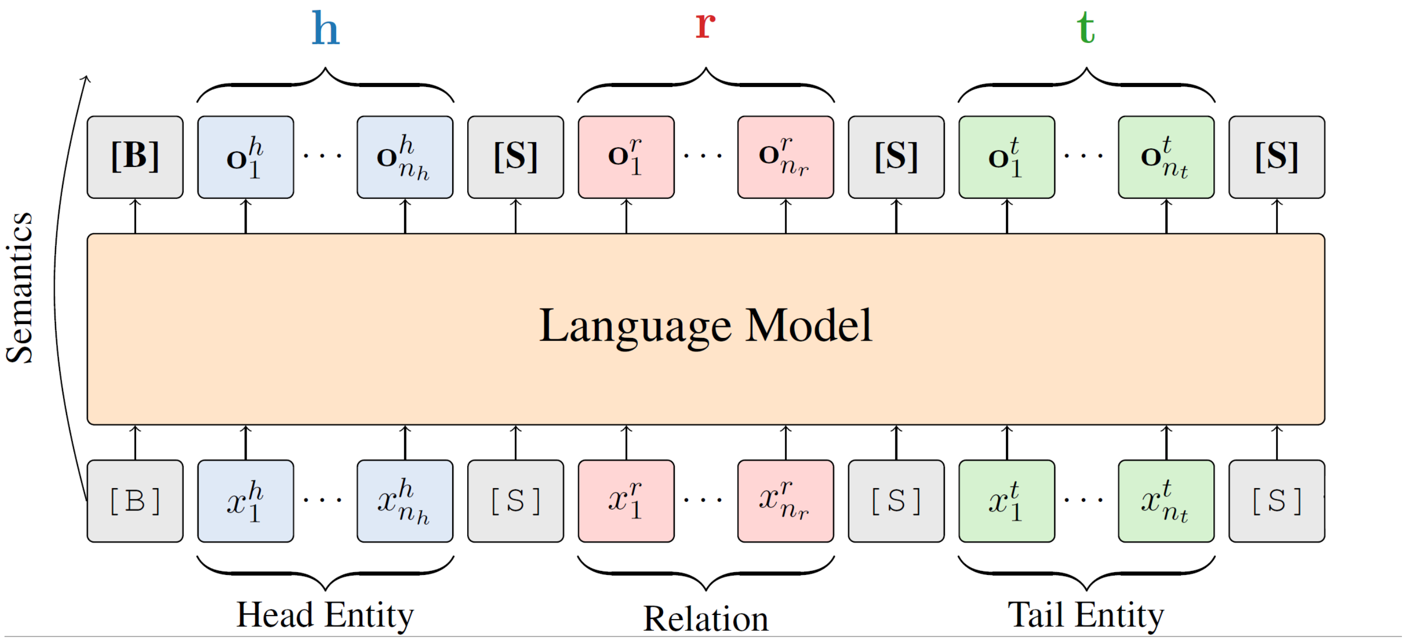


Semantic Embedding

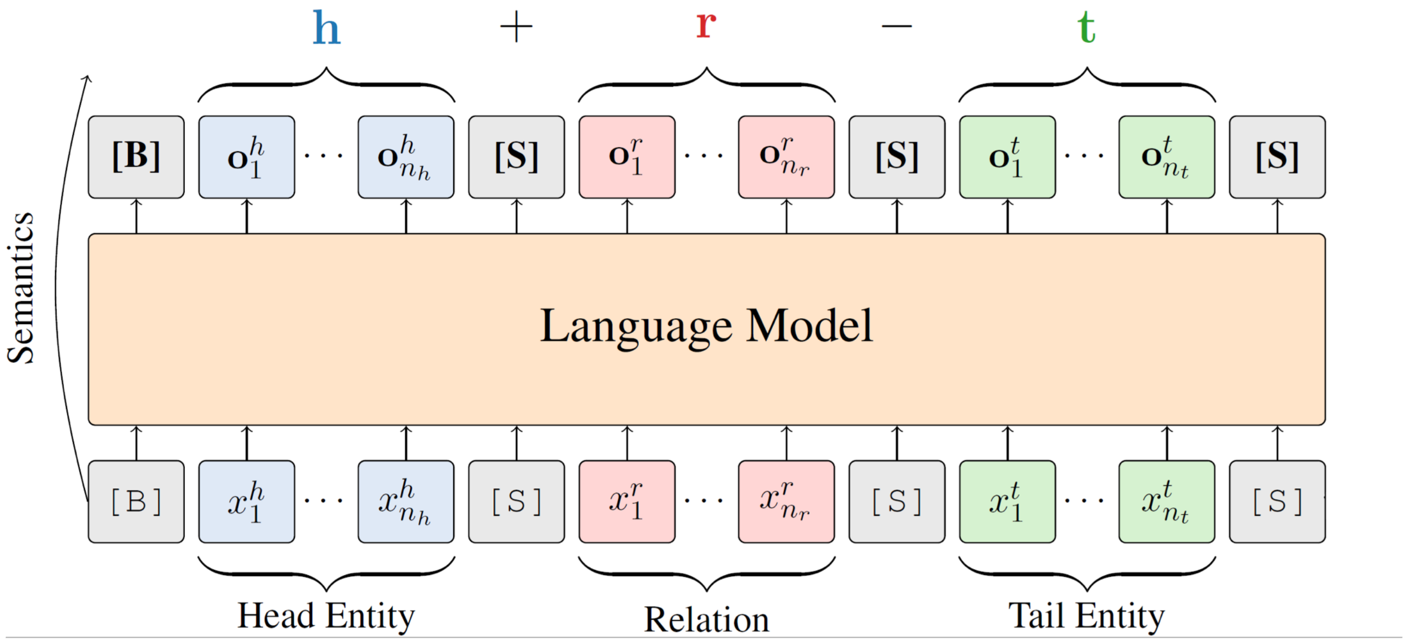


Semantic embeddings take advantage of the semantic information learned by pre-trained language models

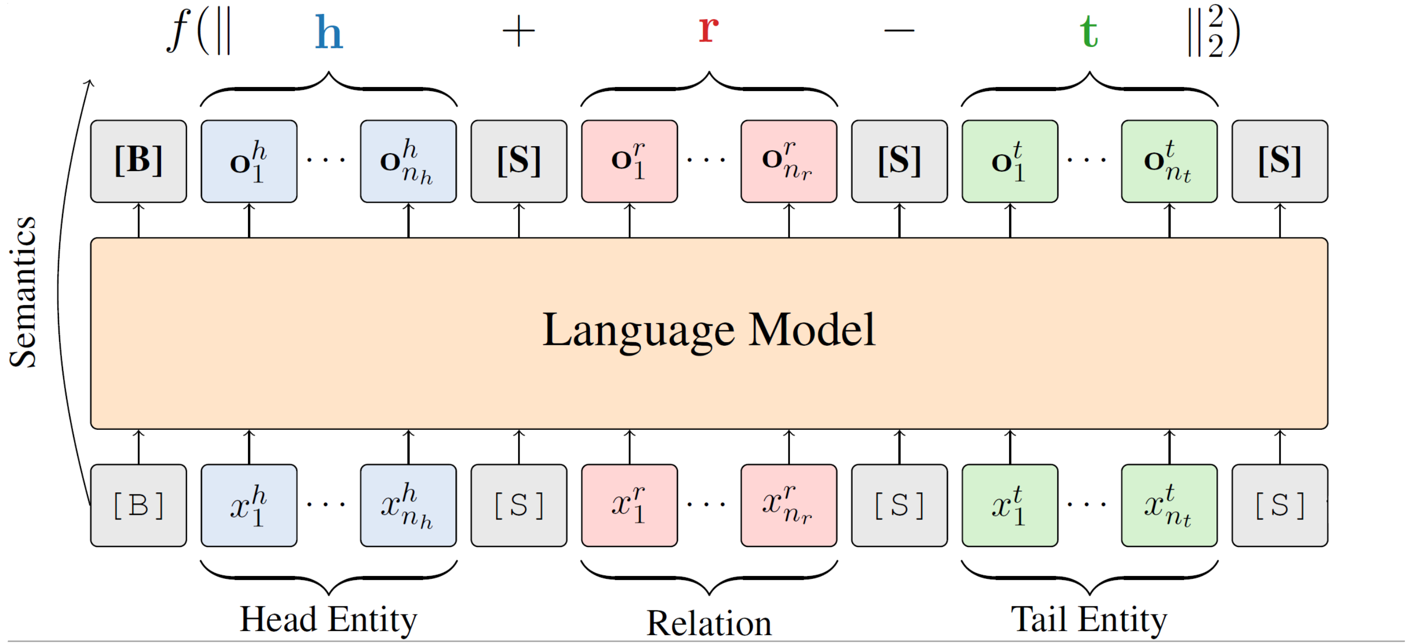
Structure Embedding



Structure Embedding

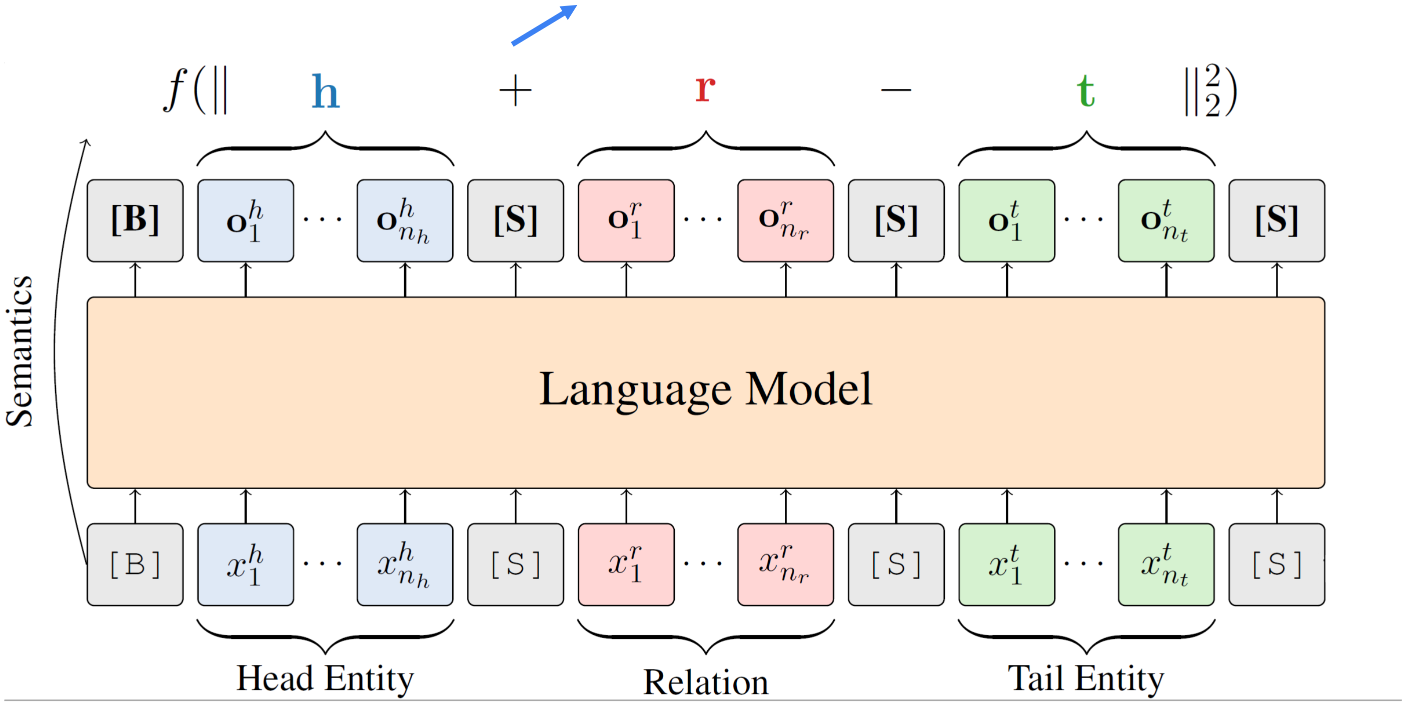


Structure Embedding



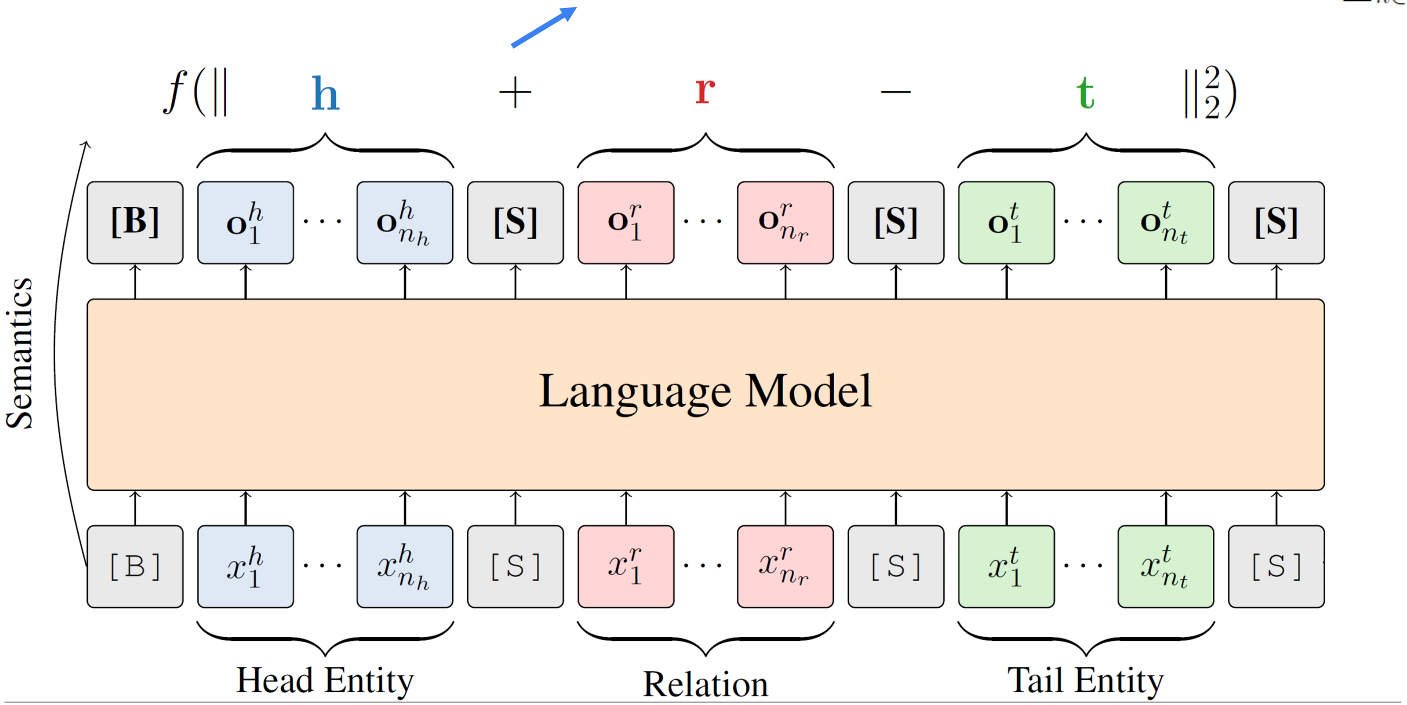
Structure Embedding

$$f(\mathbf{h}, \mathbf{r}, \mathbf{t}) = b - \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2$$



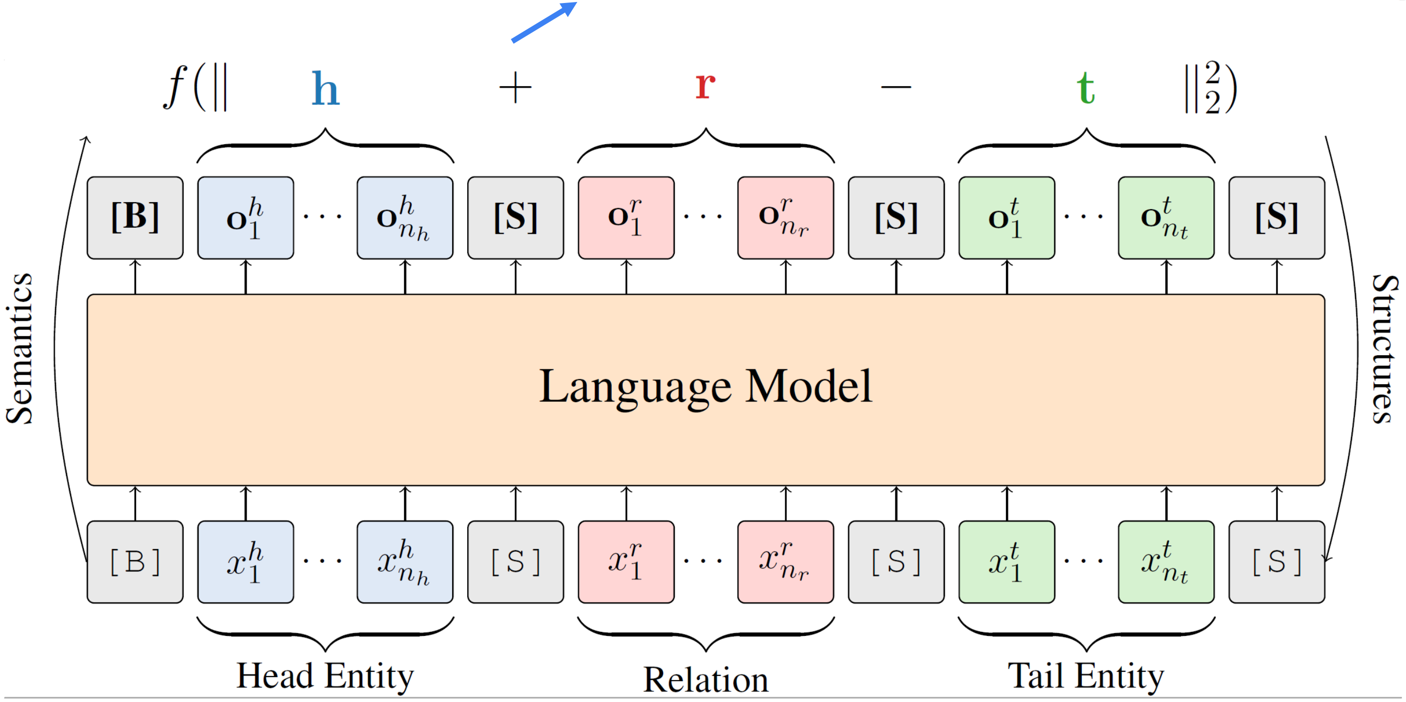
Structure Embedding

$$f(\mathbf{h},\mathbf{r},\mathbf{t}) = b - \frac{1}{2}\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_2^2 \longrightarrow \Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$



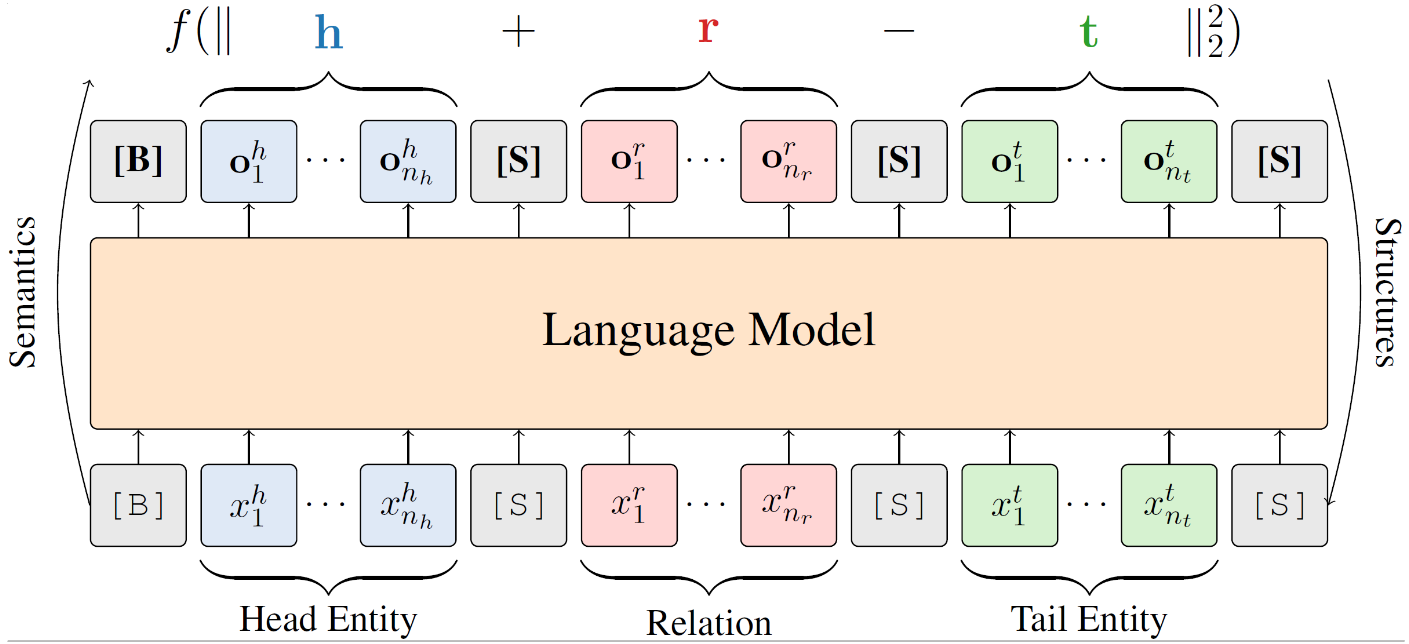
Structure Embedding

$$f(\mathbf{h},\mathbf{r},\mathbf{t}) = b - \frac{1}{2}\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_2^2 \longrightarrow \Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$



Structure Embedding

$$f(\mathbf{h},\mathbf{r},\mathbf{t}) = b - \frac{1}{2}\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|_2^2 \longrightarrow \Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$



Structure embeddings preserve the structure information by making h + r close to t

Optimization

$$\Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$

Optimization

$$\Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$

Negative sampling strategy for efficiency

$$L_h = -\log \Pr(1|h,r,t) - \sum_{i}^{n_{\text{ns}}} \mathbb{E}_{\tilde{h}_i \sim E \backslash \{h\}} \log \Pr(0|\tilde{h}_i, r, t)$$

$$\Pr(1|h,r,t) = \sigma(f(\mathbf{h},\mathbf{r},\mathbf{t}))$$

Optimization

$$\Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$

Negative sampling strategy for efficiency

Negative corrupted head

$$L_h = -\log\Pr(1|h,r,t) - \sum_{i}^{n_{\text{ns}}} \mathbb{E}_{\tilde{h}_i\sim E\setminus\{h\}} \log\Pr(0|\tilde{h}_i,r,t)$$

$$\Pr(1|h,r,t) = \sigma(f(\mathbf{h},\mathbf{r},\mathbf{t}))$$

Optimization

$$\Pr(h|r,t) = \frac{\exp(f(\mathbf{h},\mathbf{r},\mathbf{t}))}{\sum_{\tilde{h}\in E}\exp(f(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t}))}$$

Negative sampling strategy for efficiency

Negative corrupted head

$$L_h = -\log\Pr(1|h,r,t) - \sum_{i}^{n_{\text{ns}}} \mathbb{E}_{\tilde{h}_i\sim E\setminus\{h\}} \log\Pr(0|\tilde{h}_i,r,t)$$

$$\Pr(1|h,r,t) = \sigma(f(\mathbf{h},\mathbf{r},\mathbf{t}))$$

Sigmoid function

Inference

- Triplet Classification

$$p(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \begin{cases} 1 & \text{if } b - \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2 \geq \sigma \\ -1 & \text{otherwise} \end{cases}$$

- Link Prediction

$$p(\mathbf{h}, \mathbf{r}) = \arg\max_{\hat{t}} b - \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \hat{\mathbf{t}}\|_2^2$$

Inference

- Triplet Classification

Score function

$$p(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \begin{cases} 1 & \text{if } b - \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2 \geq \sigma \\ -1 & \text{otherwise} \end{cases}$$

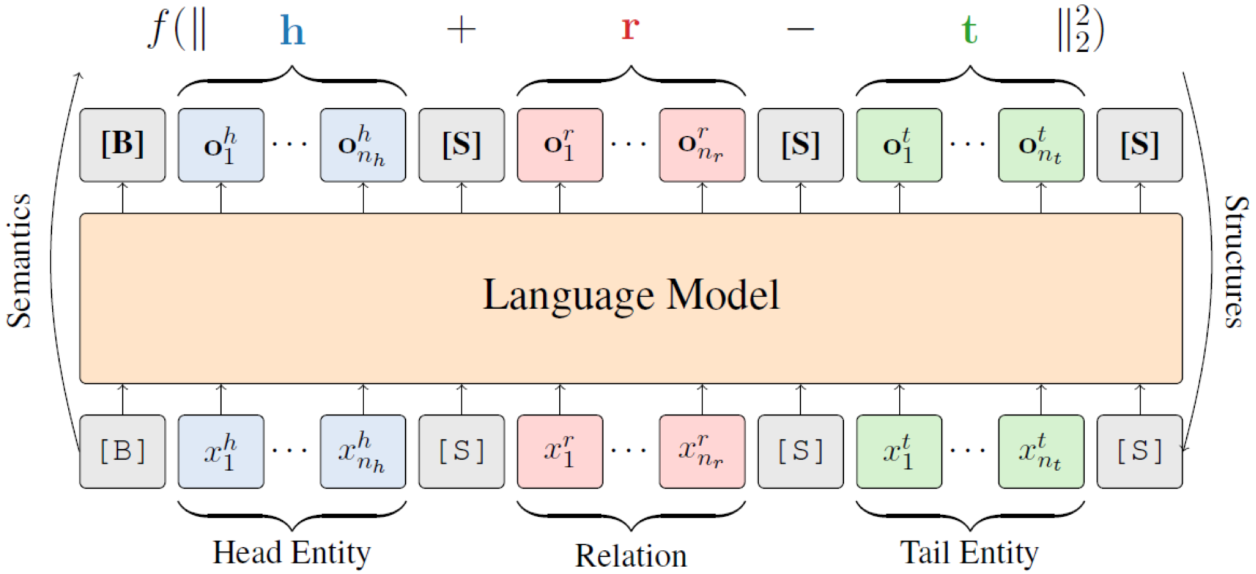
- Link Prediction

$$p(\mathbf{h}, \mathbf{r}) = \arg\max_{\hat{t}} b - \frac{1}{2}\|\mathbf{h} + \mathbf{r} - \hat{\mathbf{t}}\|_2^2$$

Score function

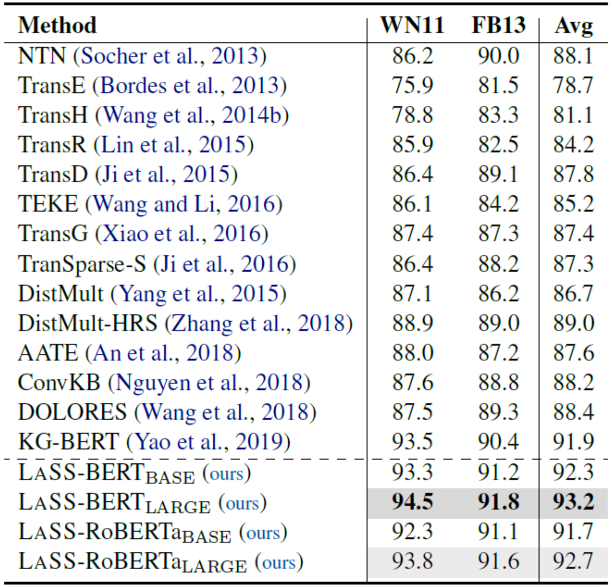
Recap of LASS

- Forward pass of LM performs semantic embedding
- Optimization of structured loss conducts structure embedding



Experiments

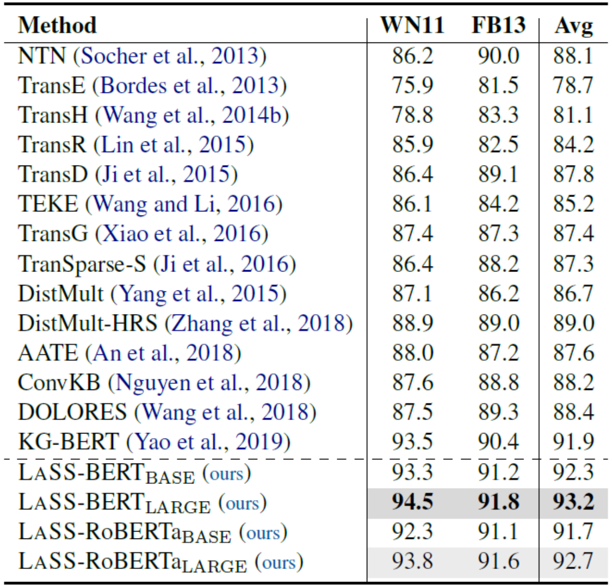
Experiments



Method	WN11	FB13	Avg
NTN (Socher et al., 2013)	86.2	90.0	88.1
TransE (Bordes et al., 2013)	75.9	81.5	78.7
TransH (Wang et al., 2014b)	78.8	83.3	81.1
TransR (Lin et al., 2015)	85.9	82.5	84.2
TransD (Ji et al., 2015)	86.4	89.1	87.8
TEKE (Wang and Li, 2016)	86.1	84.2	85.2
TransG (Xiao et al., 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al., 2016)	86.4	88.2	87.3
DistMult (Yang et al., 2015)	87.1	86.2	86.7
DistMult-HRS (Zhang et al., 2018)	88.9	89.0	89.0
AATE (An et al., 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
DOLORES (Wang et al., 2018)	87.5	89.3	88.4
KG-BERT (Yao et al., 2019)	93.5	90.4	91.9
LaSS-BERT$_{\text{BASE}}$ (ours)	93.3	91.2	92.3
LaSS-BERT$_{\text{LARGE}}$ (ours)	**94.5**	**91.8**	**93.2**
LaSS-RoBERTa$_{\text{BASE}}$ (ours)	92.3	91.1	91.7
LaSS-RoBERTa$_{\text{LARGE}}$ (ours)	93.8	91.6	92.7

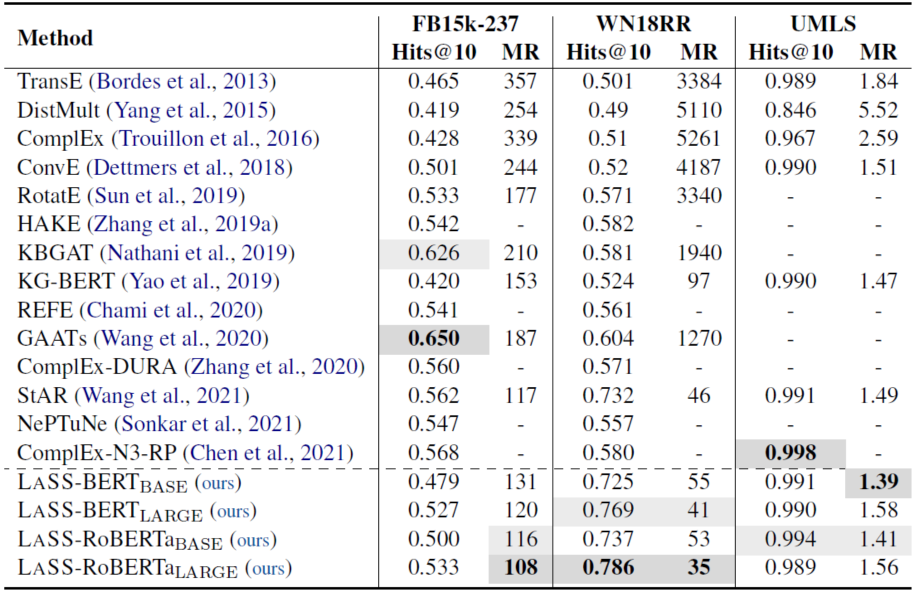
Triplet Classification

Experiments



Method	WN11	FB13	Avg
NTN (Socher et al., 2013)	86.2	90.0	88.1
TransE (Bordes et al., 2013)	75.9	81.5	78.7
TransH (Wang et al., 2014b)	78.8	83.3	81.1
TransR (Lin et al., 2015)	85.9	82.5	84.2
TransD (Ji et al., 2015)	86.4	89.1	87.8
TEKE (Wang and Li, 2016)	86.1	84.2	85.2
TransG (Xiao et al., 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al., 2016)	86.4	88.2	87.3
DistMult (Yang et al., 2015)	87.1	86.2	86.7
DistMult-HRS (Zhang et al., 2018)	88.9	89.0	89.0
AATE (An et al., 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
DOLORES (Wang et al., 2018)	87.5	89.3	88.4
KG-BERT (Yao et al., 2019)	93.5	90.4	91.9
LaSS-BERT$_{\text{BASE}}$ (ours)	93.3	91.2	92.3
LaSS-BERT$_{\text{LARGE}}$ (ours)	**94.5**	**91.8**	**93.2**
LaSS-RoBERTa$_{\text{BASE}}$ (ours)	92.3	91.1	91.7
LaSS-RoBERTa$_{\text{LARGE}}$ (ours)	93.8	91.6	92.7

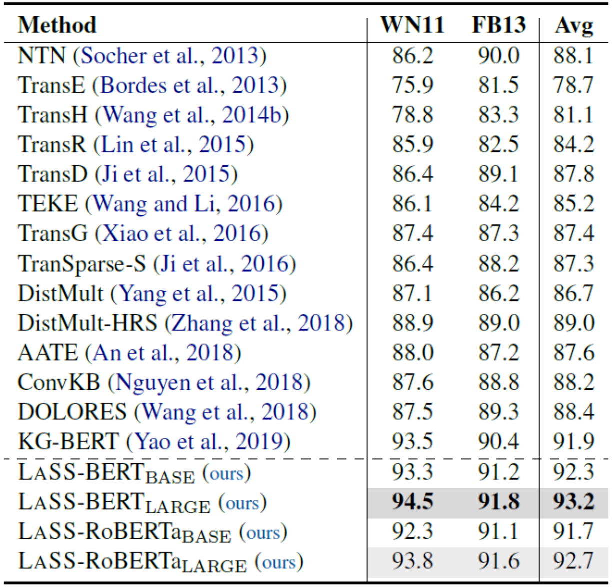
Triplet Classification



Method	FB15k-237		WN18RR		UMLS	
	Hits@10	MR	Hits@10	MR	Hits@10	MR
TransE (Bordes et al., 2013)	0.465	357	0.501	3384	0.989	1.84
DistMult (Yang et al., 2015)	0.419	254	0.49	5110	0.846	5.52
ComplEx (Trouillon et al., 2016)	0.428	339	0.51	5261	0.967	2.59
ConvE (Dettmers et al., 2018)	0.501	244	0.52	4187	0.990	1.51
RotatE (Sun et al., 2019)	0.533	177	0.571	3340	-	-
HAKE (Zhang et al., 2019a)	0.542	-	0.582	-	-	-
KBGAT (Nathani et al., 2019)	0.626	210	0.581	1940	-	-
KG-BERT (Yao et al., 2019)	0.420	153	0.524	97	0.990	1.47
REFE (Chami et al., 2020)	0.541	-	0.561	-	-	-
GAATs (Wang et al., 2020)	**0.650**	187	0.604	1270	-	-
ComplEx-DURA (Zhang et al., 2020)	0.560	-	0.571	-	-	-
StAR (Wang et al., 2021)	0.562	117	0.732	46	0.991	1.49
NePTuNe (Sonkar et al., 2021)	0.547	-	0.557	-	-	-
ComplEx-N3-RP (Chen et al., 2021)	0.568	-	0.580	-	**0.998**	-
LaSS-BERT$_{\text{BASE}}$ (ours)	0.479	131	0.725	55	0.991	**1.39**
LaSS-BERT$_{\text{LARGE}}$ (ours)	0.527	120	0.769	41	0.990	1.58
LaSS-RoBERTa$_{\text{BASE}}$ (ours)	0.500	116	0.737	53	0.994	1.41
LaSS-RoBERTa$_{\text{LARGE}}$ (ours)	0.533	**108**	**0.786**	**35**	0.989	1.56

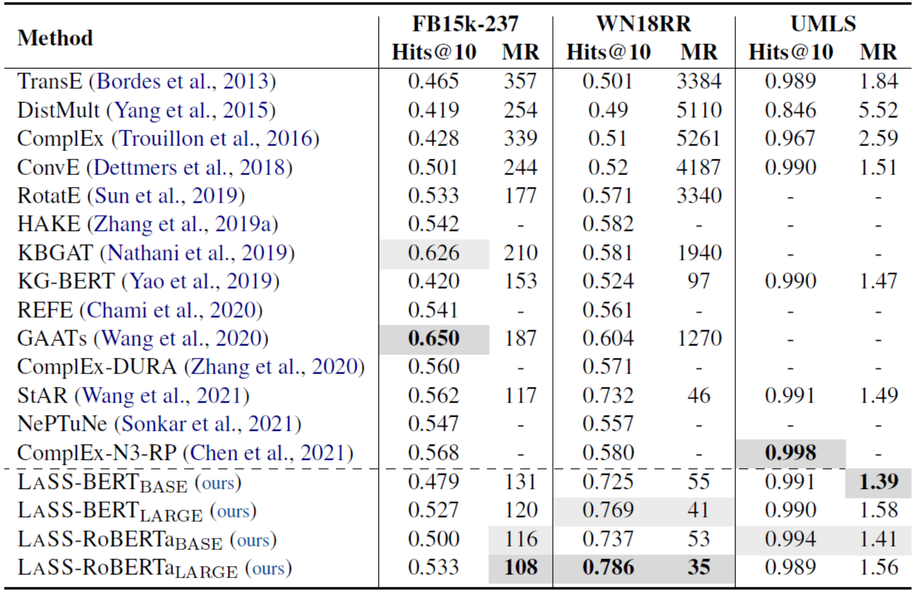
Link Prediction

Experiments



Method	WN11	FB13	Avg
NTN (Socher et al., 2013)	86.2	90.0	88.1
TransE (Bordes et al., 2013)	75.9	81.5	78.7
TransH (Wang et al., 2014b)	78.8	83.3	81.1
TransR (Lin et al., 2015)	85.9	82.5	84.2
TransD (Ji et al., 2015)	86.4	89.1	87.8
TEKE (Wang and Li, 2016)	86.1	84.2	85.2
TransG (Xiao et al., 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al., 2016)	86.4	88.2	87.3
DistMult (Yang et al., 2015)	87.1	86.2	86.7
DistMult-HRS (Zhang et al., 2018)	88.9	89.0	89.0
AATE (An et al., 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al., 2018)	87.6	88.8	88.2
DOLORES (Wang et al., 2018)	87.5	89.3	88.4
KG-BERT (Yao et al., 2019)	93.5	90.4	91.9
LaSS-BERT$_{\text{BASE}}$ (ours)	93.3	91.2	92.3
LaSS-BERT$_{\text{LARGE}}$ (ours)	**94.5**	**91.8**	**93.2**
LaSS-RoBERTa$_{\text{BASE}}$ (ours)	92.3	91.1	91.7
LaSS-RoBERTa$_{\text{LARGE}}$ (ours)	93.8	91.6	92.7

Triplet Classification

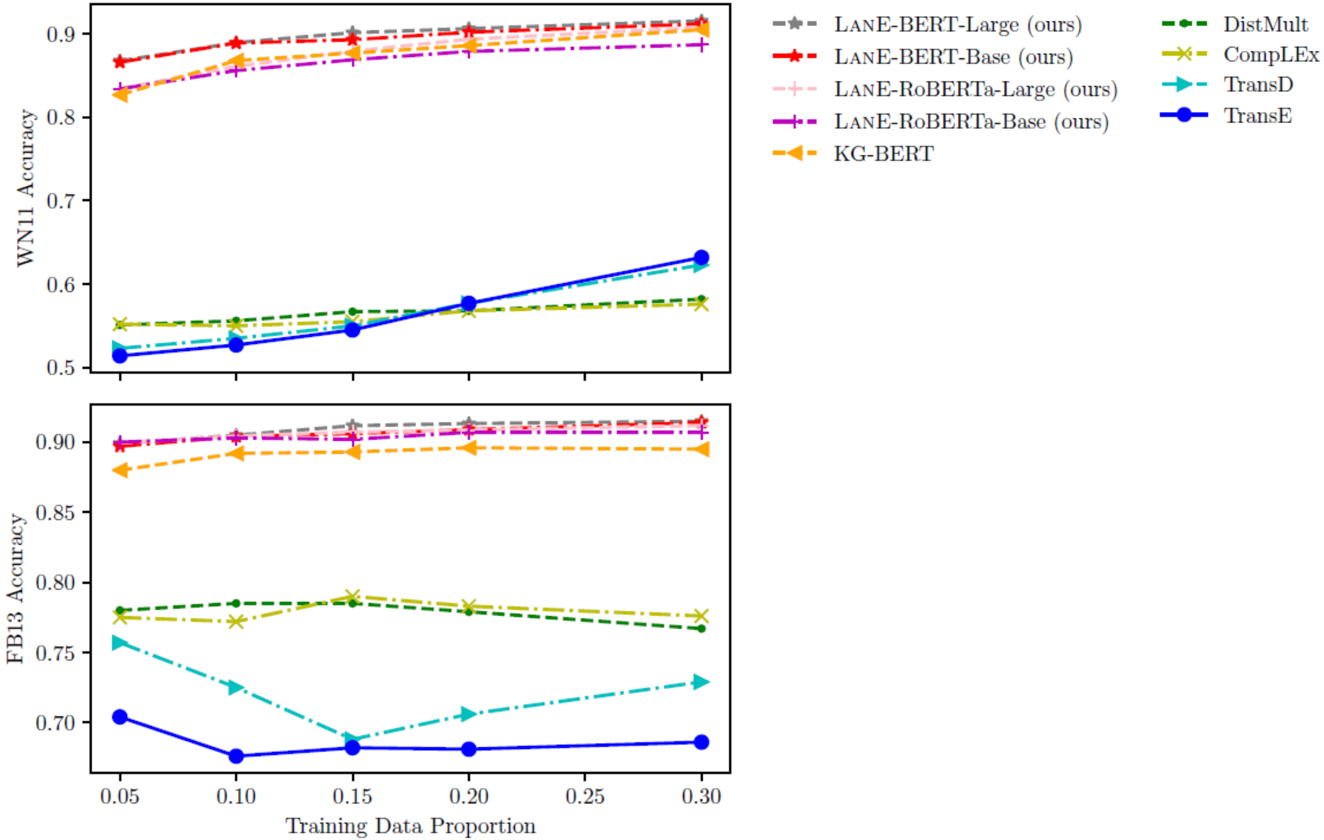


Method	FB15k-237		WN18RR		UMLS	
	Hits@10	MR	Hits@10	MR	Hits@10	MR
TransE (Bordes et al., 2013)	0.465	357	0.501	3384	0.989	1.84
DistMult (Yang et al., 2015)	0.419	254	0.49	5110	0.846	5.52
ComplEx (Trouillon et al., 2016)	0.428	339	0.51	5261	0.967	2.59
ConvE (Dettmers et al., 2018)	0.501	244	0.52	4187	0.990	1.51
RotatE (Sun et al., 2019)	0.533	177	0.571	3340	-	-
HAKE (Zhang et al., 2019a)	0.542	-	0.582	-	-	-
KBGAT (Nathani et al., 2019)	0.626	210	0.581	1940	-	-
KG-BERT (Yao et al., 2019)	0.420	153	0.524	97	0.990	1.47
REFE (Chami et al., 2020)	0.541	-	0.561	-	-	-
GAATs (Wang et al., 2020)	**0.650**	187	0.604	1270	-	-
ComplEx-DURA (Zhang et al., 2020)	0.560	-	0.571	-	-	-
StAR (Wang et al., 2021)	0.562	117	0.732	46	0.991	1.49
NePTuNe (Sonkar et al., 2021)	0.547	-	0.557	-	-	-
ComplEx-N3-RP (Chen et al., 2021)	0.568	-	0.580	-	**0.998**	-
LaSS-BERT$_{\text{BASE}}$ (ours)	0.479	131	0.725	55	0.991	**1.39**
LaSS-BERT$_{\text{LARGE}}$ (ours)	0.527	120	0.769	41	0.990	1.58
LaSS-RoBERTa$_{\text{BASE}}$ (ours)	0.500	116	0.737	53	0.994	1.41
LaSS-RoBERTa$_{\text{LARGE}}$ (ours)	0.533	**108**	**0.786**	**35**	0.989	1.56

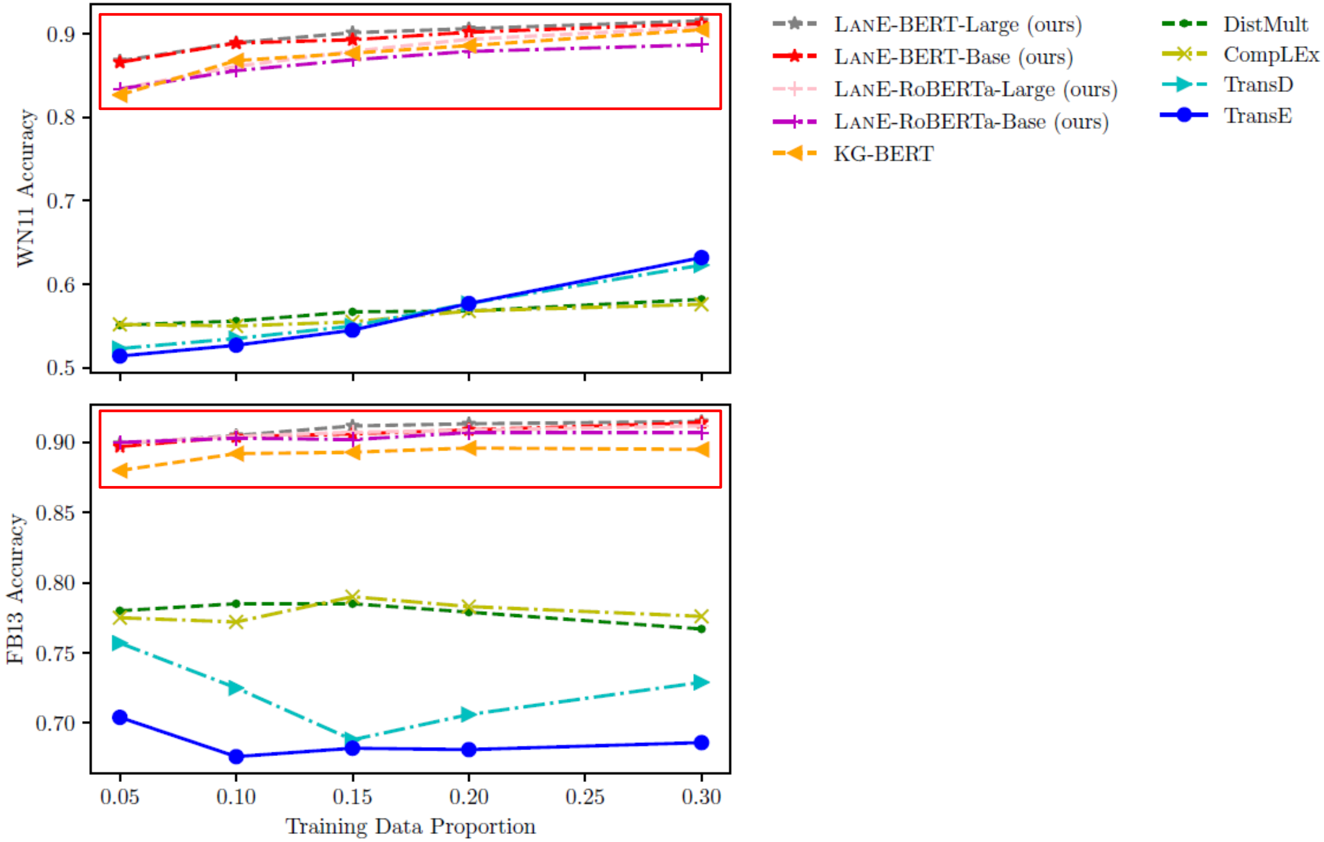
Link Prediction

Our approach achieves state-of-the-art performance on knowledge graph completion

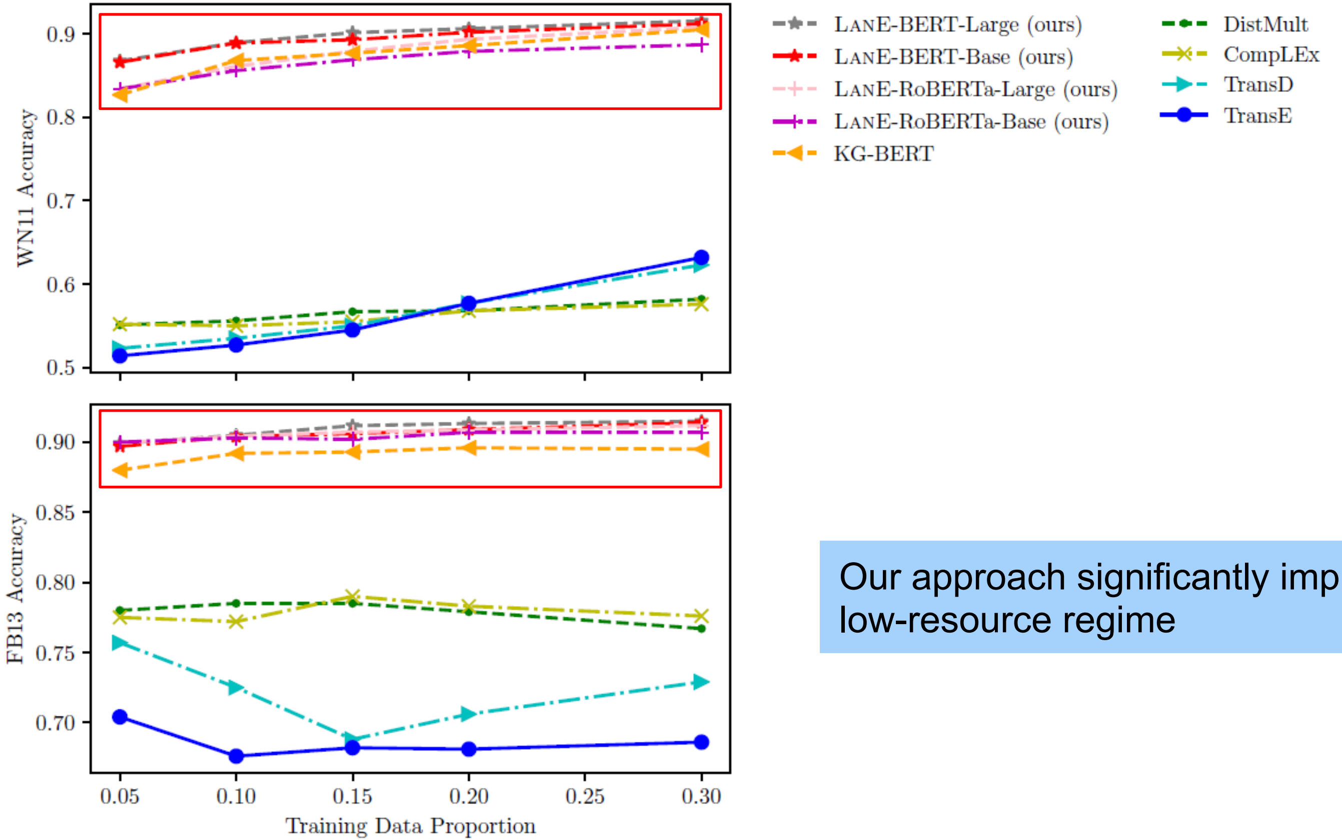
Low-Resource Settings



Low-Resource Settings

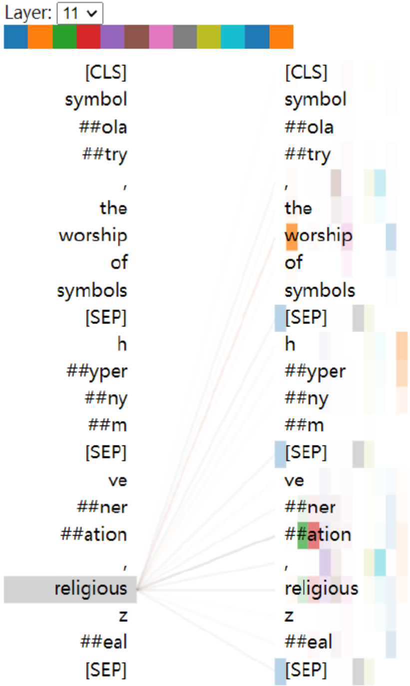


Low-Resource Settings

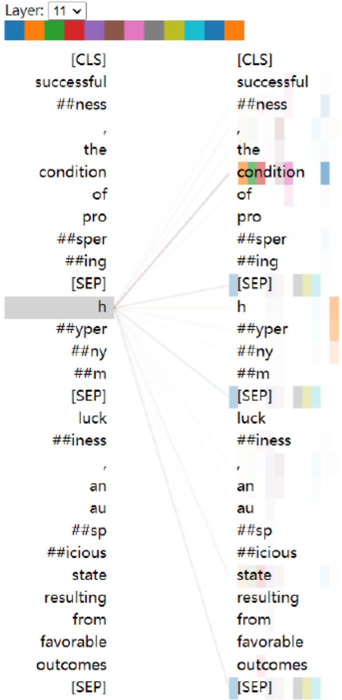


Our approach significantly improves performance in a low-resource regime

Case Study

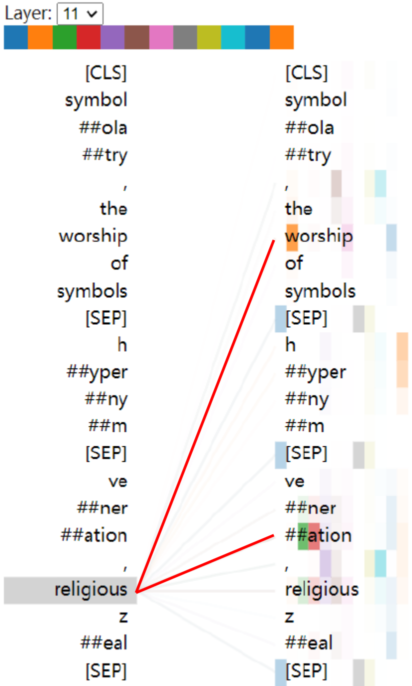


(a) Semantics.

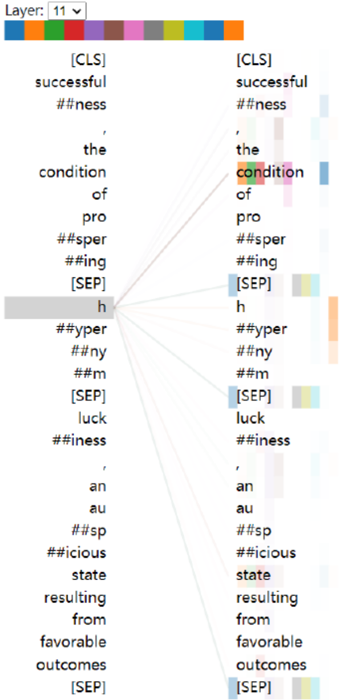


(b) Structures.

Case Study

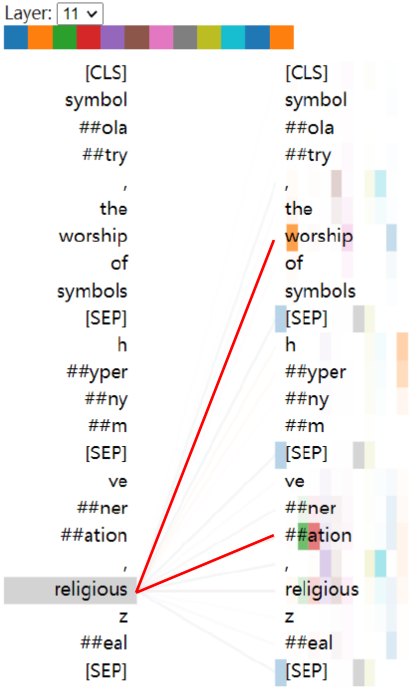

Layer: 11

(a) Semantics.


Layer: 11

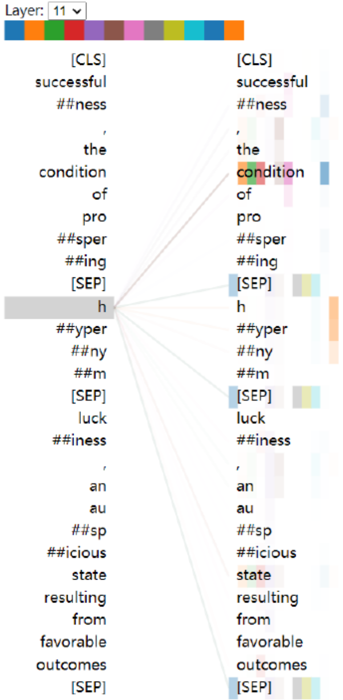
(b) Structures.

Case Study



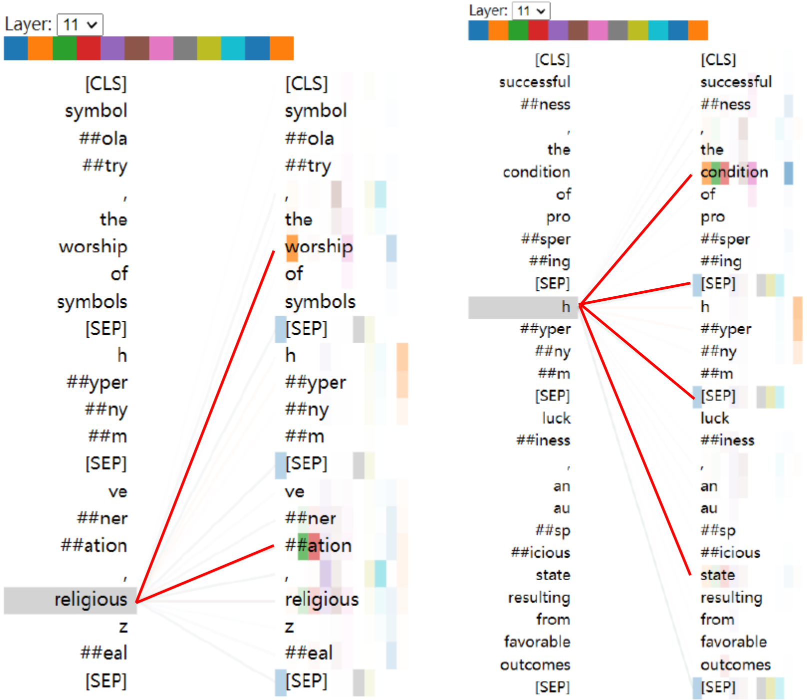
“religious” attends to
“worship and “veneration”

(a) Semantics.



(b) Structures.

Case Study

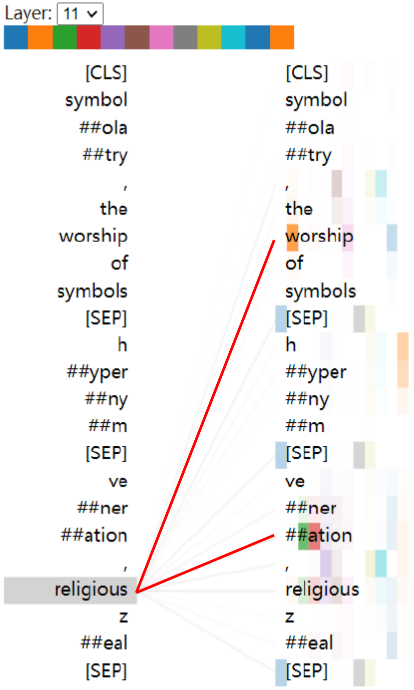


(a) Semantics.

(b) Structures.

“religious” attends to “worship and “veneration”

Case Study



Layer: 11

(a) Semantics.

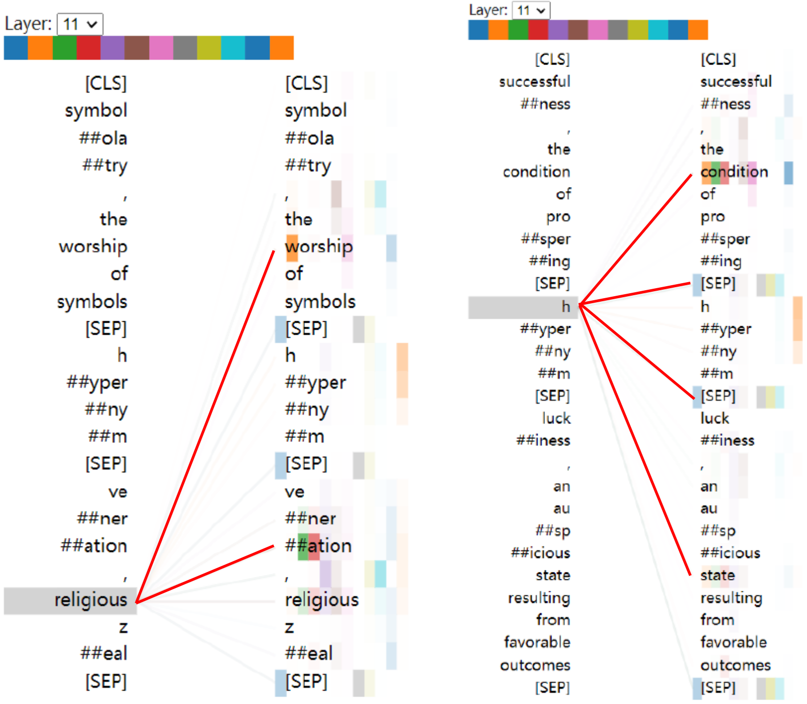
"religious" attends to "worship and "veneration"

Layer: 11

(b) Structures.

"hypernym" attends to "condition", "state" and "[SEP]"

Case Study



(a) Semantics. (b) Structures.

“religious” attends to “worship and “veneration”

“hypernym” attends to “condition”, “state” and “[SEP]”

LASS captures both semantics and structures of the triplet

Conclusion

Conclusion

LASS

Conclusion

A new embedding method that leverages both semantics and structures for KG completion

LASS

Conclusion

A new embedding method that leverages both semantics and structures for KG completion

LASS

The forward pass captures semantics and the loss reconstructs structures

Conclusion

A new embedding method that leverages both semantics and structures for KG completion

Achieves state-of-the-art performance on KG completion

LASS

The forward pass captures semantics and the loss reconstructs structures

Conclusion

A new embedding method that leverages both semantics and structures for KG completion

Achieves state-of-the-art performance on KG completion

LASS

The forward pass captures semantics and the loss reconstructs structures

Significantly improves performance in a low-resource regime

Thank you for your time!

Code: https://github.com/pkusjh/LASS