

基于准确文本增强的知识表示学习方法

Accurate Text-Enhanced Knowledge Graph Representation Learning.

In: NAACL2018.

安波, 韩先培, 孙乐

联系人: 韩先培

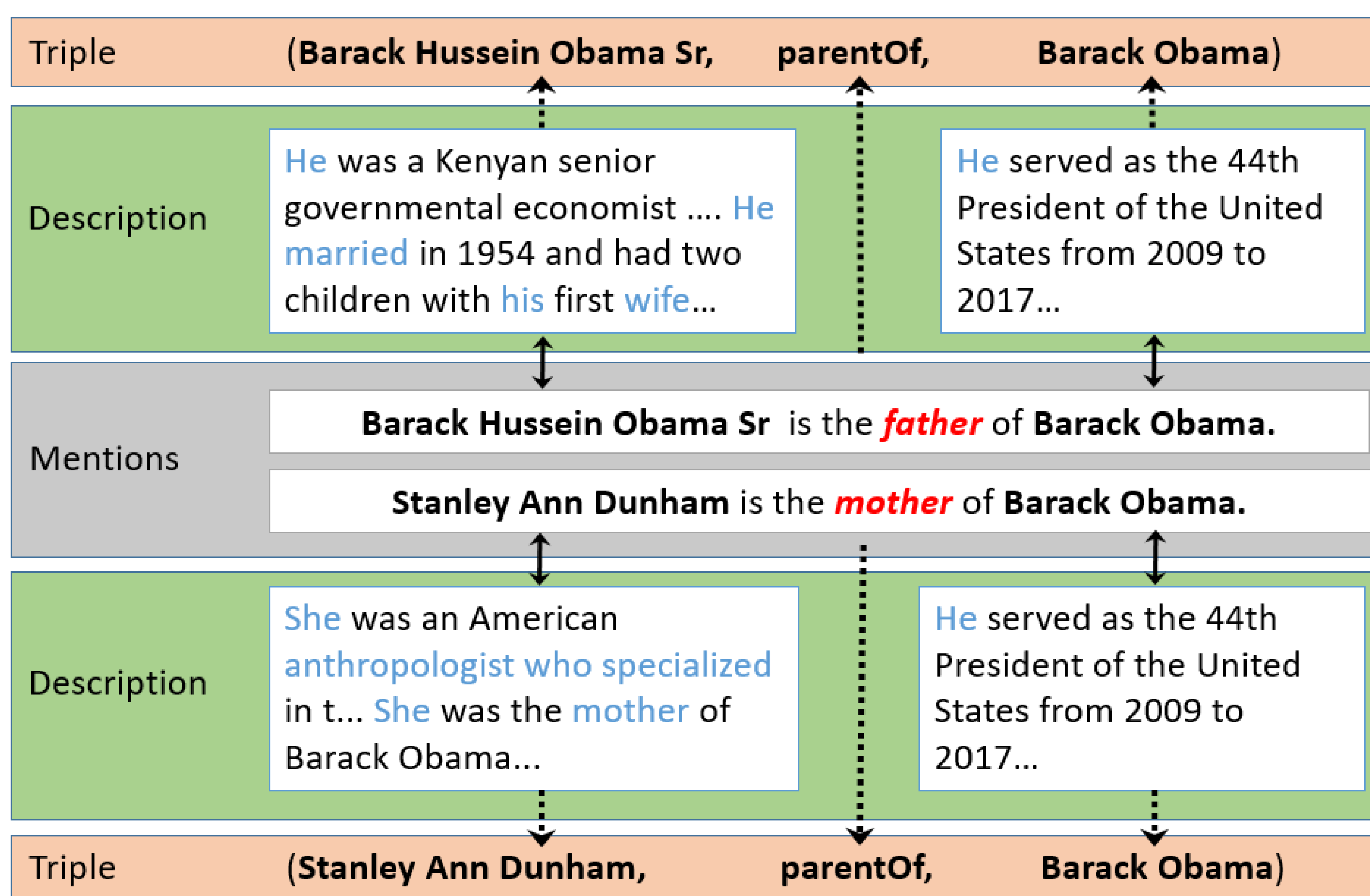
联系方式: xianpei@iscas.ac.cn

背景

- 知识表示学习方法通过将知识图谱中的实体和关系映射到向量表示取得了广泛的应用
- 传统方法通过文本与实体对的共现信息来增强实体和关系的表示, **忽略了知识的歧义性**
 - 同一关系在不同三元组中表达的语义不同
 - 不同关系关于实体的不同属性

准确文本增强的知识表示学习

- 文本信息是对知识的有效补充, 可以明确关系的语义, 同时包含实体的不同属性信息
- 抽取能够**准确描述三元组语义的文本提及**增强关系表示
- 通过**互注意力机制**来选择更好的实体属性



文本提及抽取

- 实体链接:
 - 基于字符匹配等方法实现实体与文本的匹配
- 文本提及抽取:
 - 所有包含实体对的句子作为候选
 - 基于语义相似度的过滤低质量提及

Triple: (Barack Obama, /people/person/nationality, USA)

m_1 : Barack and Michelle have leased a house in America.

m_2 : In 1961 Obama was born in Hawaii, US. ✓

m_3 : Obama announced that the United States combat mission in Iraq was over.

文本增强的知识表示学习

- 文本表示: 基于BiLSTM+Mutual Attention的文本表示;

$$a_i(e) = \frac{\exp(\text{score}(\vec{h}_i, \vec{r}^j))}{\sum_{i'} \exp(\text{score}(\vec{h}_{i'}, \vec{r}^j))}$$

$$a_i(r) = \frac{\exp(\text{score}(\vec{h}_i, \vec{e}))}{\sum_{i'} \exp(\text{score}(\vec{h}_{i'}, \vec{e}))}$$

$$\text{score}(\vec{h}_i, \vec{r}^j) = \vec{h}_i^T W_e \vec{r}^j$$

$$\text{score}(\vec{h}_i, \vec{e}) = \vec{h}_i^T W_r \vec{e}$$

$$\vec{e}^* = \tanh(\vec{a}_e^T H_e)$$

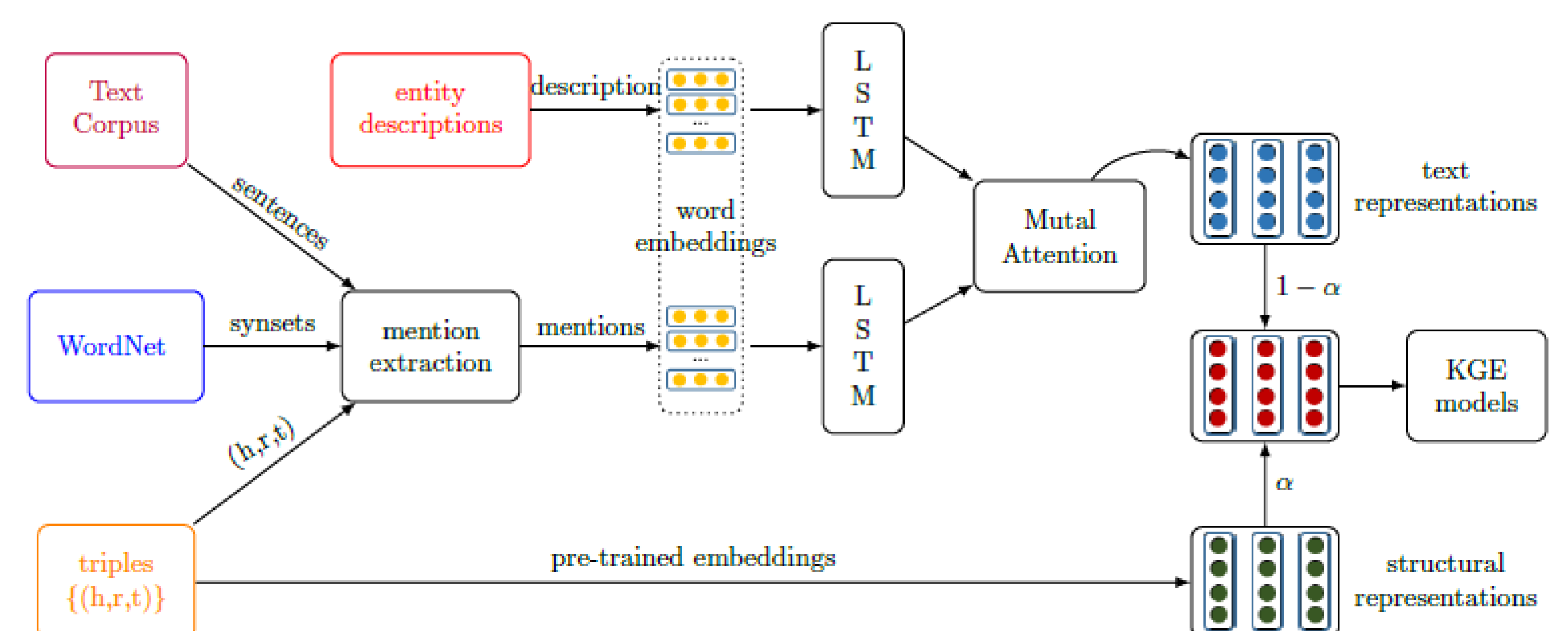
$$\vec{r}^* = \tanh(\vec{a}_r^T H_r)$$

- 使用文本表示增强实体、关系的表示;

$$Re(\vec{r}_{ate}) = \alpha \cdot Re(\vec{r}_{kg}) + (1 - \alpha) \cdot \vec{r}^*, 0 \leq \alpha \leq 1$$

$$Re(\vec{h}_{ate}) = \alpha \cdot Re(\vec{h}_{kg}) + (1 - \alpha) \cdot \vec{e}_h^*, 0 \leq \alpha \leq 1$$

$$Re(\vec{t}_{ate}) = \alpha \cdot Re(\vec{t}_{kg}) + (1 - \alpha) \cdot \vec{e}_t^*, 0 \leq \alpha \leq 1$$



实验

- 在链接预测和三元组分类任务上取得了当前最好的结果

Models	WN18		FB15K		Models	WN11	FB13	AVG.	
	MR	Hit10	MR	Hit10					
Others	UnS	304	38.2	154	40.8	TransD	86.4	89.1	87.8
	SME	533	74.1	979	6.3	TransG	87.4	87.3	87.4
	TransD	212	92.2	91	77.3	Mainfold	87.5	87.2	87.4
	TransG	345	94.7	50	88.2	TransE	75.9	81.5	78.7
	Mainfold	-	94.9	-	88.1	TEKE.E	84.1	75.1	79.6
Jointly	LSTM	95	91.6	90	69.7	ATE.E	84.3	75.4	79.9
	A-LSTM	123	90.9	73	75.5	AATE.E	86.1	86.4	86.3
TransE	TransE	251	89.2	125	47.1	TransH	78.8	83.3	81.1
	TEKE.E	127	93.8	79	67.6	TEKE.H	84.8	84.2	84.5
	ATE.E	158	91.7	89	57.1	ATE.H	85.1	83.9	84.5
	AATE.E	123	94.1	76	76.1	AATE.H	86.7	86.2	86.5
TransH	TransH	303	86.7	84	58.5	TransR	85.9	82.5	84.2
	TEKE.H	128	93.6	75	70.4	TEKE.R	86.1	81.6	83.7
	ATE.H	167	92.5	80	68.2	ATE.R	86.2	84.4	85.3
	AATE.H	132	94.0	73	74.6	AATE.R	86.4	85.2	85.8
TransR	TransR	219	91.7	78	65.5	TransR	85.9	82.5	84.2
	TEKE.R	203	92.3	79	68.5	TEKE.R	86.1	81.6	83.7
	ATE.R	210	92.1	80	67.2	ATE.R	86.2	84.4	85.3
	AATE.R	185	93.7	77	69.4	AATE.R	86.4	85.2	85.8
Complex	Complex	219	94.7	78	84.0	Complex	86.2	85.7	86.0
	ATE.C	217	94.7	61	86.2	ATE.C	87.2	87.1	87.2
	AATE.C	179	94.9	52	88.0	AATE.C	88.0	87.2	87.6

Relation Category	Prediction Head (Hits@10)			Prediction Tail (Hits@10)		
	1-to-N	N-to-1	N-to-N	1-to-N	N-to-1	N-to-N
#Triples in Test	2,078	6,084	109,526	2,078	6,084	109,526
Jointly(A-LSTM)	95.1	21.1	47.9	30.8	94.7	53.1
TransE	65.7	18.2	47.2	19.7	66.7	50.0
ATE.E	80.2	22.1	47.6	20.3	67.7	60.0
AATE.E	96.1	35.2	49.1	32.2	98.3	60.3