

Knowledge Graph and Text Jointly Embedding

이봉석

- Author

- Zhen Wang
- Jianwen Zhang
- Jianlin Feng
- Zheng Chen

- Title of Conference(Journal)

- EMNLP 2014

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \mathcal{L}_A \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

이 LOSS를 maximiz하는 방향으로 학습한다.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

(h, r, t)인 triple에 대해서 점수를 $z(h, r, t) = b - \frac{1}{2}\|h + r - t\|^2$ 를 통해 구하고 b는 수치 안정성을 위한 지정된 바이어스에 대한 상수이고 논문에서는 7이 합리적인 선택이라고 한다.

02. Modeling

$$\mathcal{L} = \boxed{\mathcal{L}_K} + \mathcal{L}_T + \mathcal{L}_A \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Knowledge Model

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

$$\Pr(h|r, t) = \frac{\exp\{z(\mathbf{h}, \mathbf{r}, \mathbf{t})\}}{\sum_{\tilde{h} \in \mathcal{I}} \exp\{z(\tilde{\mathbf{h}}, \mathbf{r}, \mathbf{t})\}}$$

$$\mathcal{L}_f(h, r, t) = \log \Pr(h|r, t) + \log \Pr(t|h, r) \\ + \log \Pr(r|h, t)$$

$$\mathcal{L}_K = \sum_{(h, r, t) \in \Delta} \mathcal{L}_f(h, r, t)$$

\mathbf{h}, \mathbf{t} 는 Vocab_emb에서
엔티티부분에서의 벡터값이고
 \mathbf{r} 은 Rel_emb에서의 벡터값이다.

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \boxed{\mathcal{L}_T} + \mathcal{L}_A \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Text Model

Skip-gram에서 윈도우안에 있는 단어 pair (w, v) 에 대해서 이 두 단어사이에는 어떤 r_{wv} 가 있다고 가정하자. 하지만 이렇게 되면 너무 많은 r_{wv} 벡터가 필요하므로 auxiliary variables를 사용하자.

$$w' = w + r_{wv}$$

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \boxed{\mathcal{L}_T} + \mathcal{L}_A$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

(9) Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Text Model

$$z(\mathbf{w}, \mathbf{r}_{wv}, \mathbf{v}) \triangleq z(\mathbf{w}', \mathbf{v}) = b - \frac{1}{2} \|\mathbf{w}' - \mathbf{v}\|^2$$

$$\Pr(w|r_{wv}, v) \triangleq \Pr(w|v) = \frac{\exp\{z(\mathbf{w}', \mathbf{v})\}}{\sum_{\tilde{w} \in \mathcal{V}} \exp\{z(\tilde{\mathbf{w}}', \mathbf{v})\}}$$

$$\mathcal{L}_T = \sum_{(w,v) \in \mathcal{C}} n_{wv} \log \Pr(w|v).$$

w 는 Words_aux_emb에의 벡터값이고
 v 는 Vocab_emb에서 단어부분의
벡터값이다.

(구현되어 있는 코드에서는 서로 반대였다.)

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \boxed{\mathcal{L}_T} + \mathcal{L}_A$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

(9) Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Text Model

$$z(\mathbf{w}, \mathbf{r}_{wv}, \mathbf{v}) \triangleq z(\mathbf{w}', \mathbf{v}) = b - \frac{1}{2} \|\mathbf{w}' - \mathbf{v}\|^2$$

$$\Pr(w|r_{wv}, v) \triangleq \Pr(w|v) = \frac{\exp\{z(\mathbf{w}', \mathbf{v})\}}{\sum_{\tilde{w} \in \mathcal{V}} \exp\{z(\tilde{\mathbf{w}}', \mathbf{v})\}}$$

$$\mathcal{L}_T = \sum_{(w,v) \in \mathcal{C}} n_{wv} \log \Pr(w|v).$$

Skip-gram에서는 두 단어의 확률을
로 정의

$$\Pr(w|v) = \frac{\exp\{\mathbf{w}'^T \mathbf{v}\}}{\sum_{\tilde{w} \in \mathcal{V}} \exp\{\tilde{\mathbf{w}}'^T \mathbf{v}\}}$$

이것은 inner product에 기반한 것이고
여기서의 모델의 계산은 거리에 기반한
것이다. 만약 각각의 w 에 대해서 $\|\mathbf{w}\| = 1$
이면 $\mathbf{w}'^T \mathbf{v} = 1 - \frac{1}{2} \|\mathbf{w}' - \mathbf{v}\|^2$ 이므로
기존의 Skip-gram모델과 동일하다.

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \boxed{\mathcal{L}_A} \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Alignment Model

앞에서 임베딩을 하면 지식 모델과 텍스트 모델이 서로 다른 space에 임베딩 되어 있다. 2개의 공간에 임베딩된 값을 하나로 합쳐주는 역할이 Alignment Model이다.

여기에는 \mathcal{L}_{AA} 방식과 \mathcal{L}_{AN} 방식이 있다.

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \boxed{\mathcal{L}_A} \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Alignment Model

\mathcal{L}_{AA} 방식은 Text Model의 데이터를 이용하여 alignment 하는 방식이다.

$$\mathcal{L}_{AA} = \sum_{(w,v) \in \mathcal{C}, v \in \mathcal{A}} \log \Pr(w|e_v)$$

Word pair (w, v)를 word-entity로 바뀌서 계산한다.

w는 Vocab_emb에서 단어부분의 벡터값이고

v는 Vocab_emb에서 엔티티부분의 벡터값이다.

02. Modeling

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \boxed{\mathcal{L}_A} \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

Vocab_emb : 엔티티와 단어를 임베딩한 공간

Rel_emb : 릴레이션을 임베딩한 공간

Words_aux_emb : auxiliary variables를 임베딩한 공간

Alignment Model

\mathcal{L}_{AN} 방식은 Knowledge Model의 데이터를 이용하여 alignment하는 방식이다.

트리플 (h, r, t) 에 대해서 h 가 V 에 있으면 (w_h, r, t) 를 만들고 t 가 V 에 있으면 (h, r, w_t) 를 만들고 둘다 있으면 (w_h, r, w_t) 를 만든다.

$$\mathcal{L}_{AN} = \sum_{(h,r,t) \in \Delta} \mathbf{I}_{[w_h \in V \wedge w_t \in V]} \cdot \mathcal{L}_f(w_h, r, w_t) +$$

$$\mathbf{I}_{[w_h \in V]} \cdot \mathcal{L}_f(w_h, r, t) + \mathbf{I}_{[w_t \in V]} \cdot \mathcal{L}_f(h, r, w_t)$$

h, t 는 Vocab_emb에서 엔티티부분에서의 벡터값이고
 w_h, w_t 는 Vocab_emb에서 단어부분에서의 벡터값이고
 r 은 Rel_emb에서의 벡터값이다.

03. Experiments

Table 5: Words Analogical Reasoning Task.

Method	Accuracy (%)			Hits@10 (%)		
	Semantic	Syntactic	Total	Semantic	Syntactic	Total
Skip-gram	71.4	69.0	70.0	90.4	89.3	89.8
Jointly (anchor)	75.3	68.3	71.2	91.5	88.9	89.9
Jointly (name)	54.5	54.2	59.0	75.8	86.5	82.1
Jointly (anchor+name)	56.5	65.7	61.9	78.1	87.6	83.6

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \mathcal{L}_A \quad (9)$$

where \mathcal{L}_A could be \mathcal{L}_{AA} or \mathcal{L}_{AN} or $\mathcal{L}_{AA} + \mathcal{L}_{AN}$.

- Jointly(anchor) $\Rightarrow \mathcal{L}_A = \mathcal{L}_{AA}$
- Jointly(name) $\Rightarrow \mathcal{L}_A = \mathcal{L}_{AN}$
- Jointly(anchor + name) $\Rightarrow \mathcal{L}_A = \mathcal{L}_{AA} + \mathcal{L}_{AN}$

감사합니다