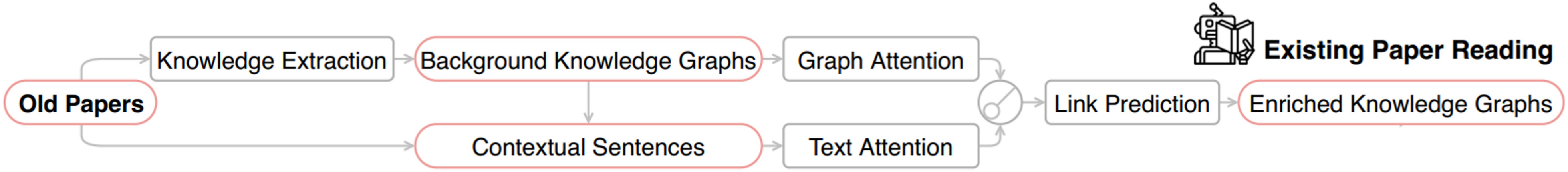
链路预测【阅读笔记】PaperRobot链路预测部分——流程梳理

2019年6月13日　10:38

链路预测部分——流程梳理



1. **大致步骤及细节公式**
   a. 随机初始化实体e和关系r等的向量表示
      i. 元组表示（头实体，关系，尾实体）

$$(e_i^h, r_i, e_i^t)$$

      ii. 针对ei实体的one-hop邻实体

$$N_{e_i} = [n_{i1}, n_{i2}, \ldots]$$

   b. 利用one-hop邻实体信息，使用self-attention和multi-head attention更新实体ei的向量表示[图结构化的实体向量表示]
      i. Self-attention

$$\boldsymbol{e}_i^{'} = \boldsymbol{W}_e \boldsymbol{e}_i, \quad \boldsymbol{n}_{ij}^{'} = \boldsymbol{W}_e \boldsymbol{n}_{ij}$$

$$c_{ij} = \mathrm{LeakyReLU}(\boldsymbol{W}_f(\boldsymbol{e}_i^{'} \oplus \boldsymbol{n}_{ij}^{'}))$$

$$\boldsymbol{c}_i^{'} = \mathrm{Softmax}(\boldsymbol{c}_i)$$

$$\boldsymbol{\epsilon}_i = \sigma\left(\sum c_{ij}^{'} \boldsymbol{n}_{ij}^{'}\right)$$

      ii. Multi-head attention

$$\tilde{\boldsymbol{e}}_i = [\boldsymbol{\epsilon}_i^0 \oplus \ldots \oplus \boldsymbol{\epsilon}_i^M]$$

c. 使用bi-LSTM将含e的语句转为编码隐藏状态H

   i. 上下文语句

$$[w_1, \ldots, w_l].$$

   ii. 经过bi-LSTM得到的隐藏状态

$$\boldsymbol{H}_s = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_l].$$

d. 使用bilinear attention得到实体e的向量表示[上下文信息的实体向量表示]

   i. Bilinear attention

$$\mu_i = \boldsymbol{e}^\top \boldsymbol{W}_s \boldsymbol{h}_i,$$
$$\boldsymbol{\mu}' = \mathrm{Softmax}(\boldsymbol{\mu})$$
$$\hat{\boldsymbol{e}} = \boldsymbol{\mu}'^\top \boldsymbol{h}_i$$

e. 使用Gated Combination结合两种实体向量表示

   i. Gate function

$$\boldsymbol{g}_e = \sigma(\tilde{\boldsymbol{g}}_e), \quad \boldsymbol{e} = \boldsymbol{g}_e \odot \tilde{\boldsymbol{e}} + (1 - \boldsymbol{g}_e) \odot \hat{\boldsymbol{e}}$$

f. 使用TransE模型训练和链路预测生成增强版的KGs

   i. Margin Loss

$$Loss = \sum_{(e_i^h, r_i, e_i^t) \in K} \sum_{(\bar{e}_i^h, \bar{r}_i, \bar{e}_i^t) \in \bar{K}} \max(0, \gamma + F(e_i^h, r_i, e_i^t) - F(\bar{e}_i^h, \bar{r}_i, \bar{e}_i^t))$$

   ii. 节点间距离分数

$$F(e_i^h, r_i, e_i^t) = \| \boldsymbol{e}_i^h + \boldsymbol{r}_i - \boldsymbol{e}_i^t \|_2^2$$

2. TansE模型[知识图谱的经典表示学习方法]——伪代码

1.

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

  $\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

  $\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

  $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

  $S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b

  $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

  for $(h, \ell, t) \in S_{batch}$ **do**

    $(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet

    $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

  end for

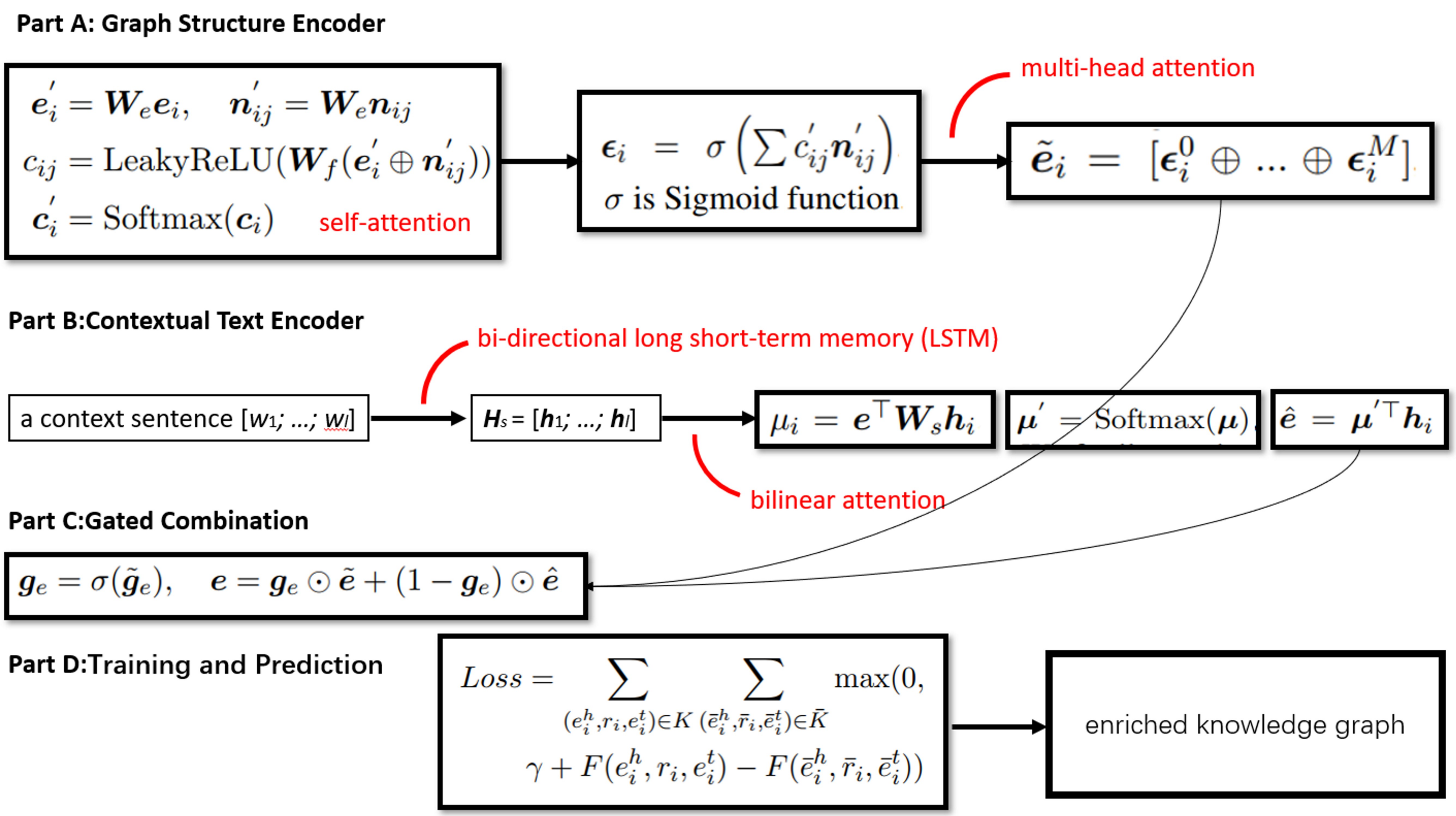
  Update embeddings w.r.t. $\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$

end loop

3. 涉及模型的超参数

Models	**Parameter**	**Value**
Link Prediction	# Multi-head	8
Multi-head hidden	8	
Entity embedding	64	
LeakyReLU α	0.2	
Margin loss γ	1	

4. 大致流程（图形式）



5. 相关疑问

- **2.3 Link Prediction第二段中**的ei is also associated with a context description si which is randomly selected from the sentences where ei occurs和**Contextual Text中**的Encoder Each entity e is also associated with a context sentence [w1; …; wl]，第二句中的context sentence是否为第一句中的si?
- **Contextual Text Encoder中**的Then we compute a bilinear attention weight for each word wi，μ=eTWshi，其e指具体的ei实体吗?
- **Gated Combination中**的where ge is an entity-dependent gate function of which each element is in [0,1], ge~ is a

learnable parameter for each entity e，其ge~的具体值/初始值为?

- **Graph Structure Encoder中**的we further perform multi-head attention on each entity, based on multiple linear transformation matrices，其multi-head attention是否为**Graph Structure Encoder中**的we perform self-attention and compute a weight distribution over Nei的self-attention的扩展?