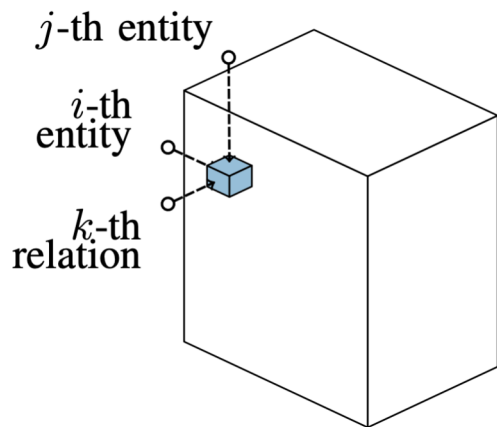
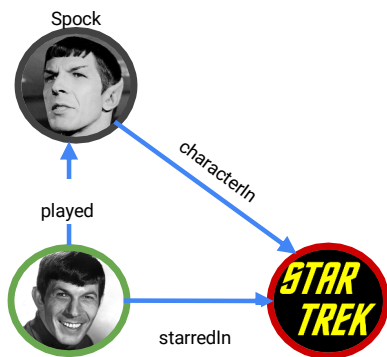


Графы знаний

Лекция 8 – Training Knowledge Graph Embeddings

М. Галкин, Д. Муромцев

TransE - Перевод ГЗ в векторное представление



Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems* (pp. 2787-2795).

Entity matrix

$$\mathbf{E} : \mathbb{R}^{|E| \times n}$$

Spock = [0.1, 0.2, 0.3]
Leonard Nimoy = [0.4, 0.8, 0.1]
Star Trek = [0.22, 0.34, 0.87]

Relations matrix

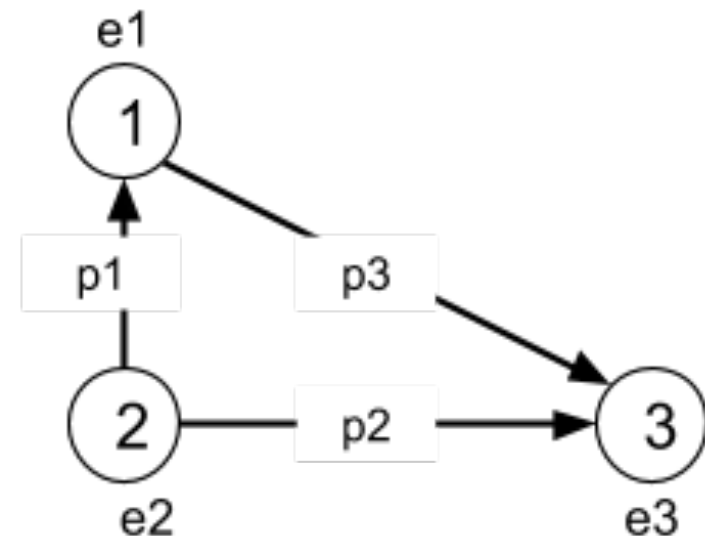
$$\mathbf{W} : \mathbb{R}^{|k| \times n}$$

characterIn = [0.1, 0.1, 0.6]
played = [0.2, 0.3, 0.4]
starredIn = [0.9, -0.2, 0.1]

TransE - Перевод ГЗ в векторное представление

При корректном триплете $\langle s \ p \ o \rangle$
предполагается, что в векторном пространстве
триплет эквивалентен векторам $s + p \approx o$,
сэмплируется противоречащий факт $s \ p \ o' | o' \neq o$.

$e_i \in E$ - множество вершин или индивидов ГЗ
 $p_j \in R$ - множество связей или отношений ГЗ
 $s, o, o' \in E$



Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems* (pp. 2787-2795).

TransE - Перевод ГЗ в векторное представление

Модель векторного представления оптимизирует функцию потерь, определяемую как

$$L = \gamma + d(s + p, o) - d(s + p, o'),$$

где функция $d(s + p, o) = s^T o + p^T (o - s)$,

а γ - гиперпараметр, задающий минимальное расстояние между векторами верного и противоречащего триплета.

Для графа, представленного выше, определены векторные представления некоторых сущностей и связей

$$e_1 = [0.1, 0.2, 0.3]$$

$$e_2 = [0.4, 0.8, 0.1]$$

$$e_3 = [0.22, 0.34, 0.87]$$

$$p_3 = [0.1, 0.1, 0.6]$$

TransE - Перевод ГЗ в векторное представление

Найдем максимальное значение гиперпараметра γ , при котором

$$L = \gamma + d(e_1 + p_3, e_3) - d(e_1 + p_3, e_2) \leq 0,6$$

$$d(e_1 + p_3, e_3) = e_1^T e_3 + p_3^T (e_3 - e_1)$$

$$d(e_1 + p_3, e_2) = e_1^T e_2 + p_3^T (e_2 - e_1)$$

$$\gamma < 0,6 - d(e_1 + p_3, e_3) + d(e_1 + p_3, e_2)$$

$$\Rightarrow \gamma < 0,6 - [e_1^T e_3 + p_3^T (e_3 - e_1)] + [e_1^T e_2 + p_3^T (e_2 - e_1)]$$

$$\Rightarrow \gamma < 0,6 + e_1^T (e_2 - e_3) + p_3^T (e_2 - e_3)$$

$$\Rightarrow \gamma < 0,6 + (e_1^T + p_3^T)(e_2 - e_3)$$

$$\Rightarrow \gamma = \gamma_{\max} \Leftrightarrow (e_1^T + p_3^T)(e_2 - e_3) = (e_1 + p_3)(e_2 - e_3)$$

TransE - Перевод ГЗ в векторное представление

Исходные вектора	e1=[0.1, 0.2, 0.3] e2=[0.4, 0.8, 0.1] e3=[0.22, 0.34, 0.87] p3=[0.1, 0.1, 0.6]		
e ₁ + p ₃	0.1 0.2 0.3	+	0.1 0.1 0.6
e ₂ – e ₃	0.4 0.22	0.8 – 0.34	0.1 0.87
(e ₁ + p ₃)(e ₂ – e ₃)	0.18*0.2 + 0.46*0.3 - 0.77*0.9 0,036 + 0,138 - 0,693		

$$L=+0,519 \leq 0.6 \rightarrow \gamma < 0.6 - 0,519 \rightarrow \gamma \leq \mathbf{0.081}$$

TransE - Перевод ГЗ в векторное представление

Algorithm 1 Learning TransE

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

```
1: initialize  $\ell \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$  for each  $\ell \in L$ 
2:            $\ell \leftarrow \ell / \|\ell\|$  for each  $\ell \in L$ 
3:            $\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$  for each entity  $e \in E$ 
4: loop
5:    $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$  for each entity  $e \in E$ 
6:    $S_{batch} \leftarrow \text{sample}(S, b)$  // sample a minibatch of size  $b$ 
7:    $T_{batch} \leftarrow \emptyset$  // initialize the set of pairs of triplets
8:   for  $(h, \ell, t) \in S_{batch}$  do
9:      $(h', \ell, t') \leftarrow \text{sample}(S'_{(h, \ell, t)})$  // sample a corrupted triplet
10:     $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$ 
11:   end for
12:   Update embeddings w.r.t. 
$$\sum_{((h, \ell, t), (h', \ell, t')) \in T_{batch}} \nabla [\gamma + d(\mathbf{h} + \ell, \mathbf{t}) - d(\mathbf{h}' + \ell, \mathbf{t}')]_+$$

13: end loop
```

Очень полезные ссылки

Еще раз о TransE с красивой визуализацией <http://pyvandenbussche.info/2017/translating-embeddings-transe/>

Тutorial по графовым эмбедингам <https://medium.com/@alaeddineayadi/hands-on-tutorial-knowledge-graph-construction-reasoning-a0524a1676c9>

OpenKE documentation <http://139.129.163.161/static/index.html>

Тutorial AmpliGraph
<https://colab.research.google.com/drive/1rylqOnm992AdP9z1aW8metlKpPuBTRGD#scrollTo=hTkVsbcKegOD>

И документация к нему <https://docs.ampligraph.org/en/1.1.0/tutorials/AmpliGraphBasicsTutorial.html>