Computer Science Department, University of Bonn

Knowledge Graph Analysis

Solutions to Exercise Sheet 8

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann

January 24, 2019

1 In Class

1. **Latent Distant Models**

   a) Distance metrics:

   Manhattan / L_1 Norm: $||\vec{x}||_1 = \sum_{i=1}^{n} |x_i|$
   Manhattan distance: $||\vec{x} - \vec{y}||_1 = \sum_{i=1}^{n} |x_i - y_i|$.

   Euclidean / L_2 Norm: $||\vec{x}||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$
   Euclidean distance: $||\vec{x} - \vec{y}||_2 = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$.

   arccos distance: $arccos(\frac{\langle \vec{x}\vec{y} \rangle}{||\vec{x}||.||\vec{y}||})$.

   b) The matrices $\vec{W}_k^s$ and $\vec{W}_k^o$ transform the global latent features $\vec{a}_i$ and $\vec{a}_j$ of the entities to representations specific for the k-th relation. Pairs of entities in an existing relationship are closer than entities in non-existing entities. There are different matrices for subject and object position of the entities for asymmetric relations. See `http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3659/3898` page 303 for details.

   For TransE, we have a relation specific offset $\vec{r}_k$, i.e. only as many parameters as latent entity features instead of two matrices per relation.

c)

$$
\begin{aligned}
f_{ijk}^{\text{TransE}} = -d(\vec{a}_i + \vec{r}_k, \vec{a}_j) &= -\|a_i + r_k - a_j\|_2^2 \\
&= -(\|a_i + r_k\|_2^2 - 2(a_i + r_k)^T a_j + \|a_j\|_2^2) \\
&= -(\|a_i\|_2^2 + 2a_i^T r_k + \|r_k\|_2^2 - 2a_i^T a_j - 2r_k^T a_j + \|a_j\|_2^2) \\
&= -(2r_k^T (a_i - a_j) - 2a_i^T a_j + \|r_k\|_2^2 + \|a_i\|_2^2 + \|a_j\|_2^2)
\end{aligned}
$$

Since we are trying to find a scoring function for TransE, we can ignore $-(\|a_i\|_2^2 + \|a_j\|_2^2) = -2$ because a_i and a_j are unit vectors. We are mainly interested in the change/improvement in the score and not the absolute value.

2. **Graph theory**

   a) $\{4, 3, 6\}$

   b) $\{(2,3), (3,5), (5,6)\}$, and $\{(2,1), (1,4), (4,5), (5,6)\}$

   c) Choosing paths uniformly and at random, $p((1,4)) \times p((4,5)) \times p((5,6)) = \frac{1}{2} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{18}$

3. **Graph Feature Models**

   a) In the common neighbour index, nodes sharing similar neighbours get a higher score. It only considers the intersection of the neighbour sets.

   In the Adamic/Ajar index, the rarer neighbors are weighted more heavily, assuming that the rarer features are more discriminating (e.g. in the case of TF-IDF based document similarity classification, phrases like "for example" are less discriminating than "logistic regression").

   The Katz index is a global index which is based on the sum of all paths, with an exponential length factor to weigh shorter paths more. Informally, according to the Katz index, the more paths there are between two vertices and the shorter the paths are, the more similar the vertices are.

   b) Common neuighbours and Adamic/Adar index are local indices as they just consider the neighbours. Katz is a global index as it considers all paths between nodes.

   c)

$$
parents(x, a) \land education(a, b) \land institution(b, c) \implies college(p, c)
$$

$$
placeOfBirth(x, a) \land peopleBornHere(a, b) \land education(b, c) \implies college(p, c)
$$

$$
\begin{aligned}
profession(x, a) \land peopleWithProfession(a, b) \land education(b, d) \\
\land institution(d, c) \implies college(p, c)
\end{aligned}
$$

   Note, that in the lecture we use to write $(subject, predicate, object)$ to denote a triple/fact. Here we use the standard notation $predicate(subject, object)$ instead.