Computer Science Department, University of Bonn

Knowledge Graph Analysis

Solutions to Exercise Sheet 10

Dr. Asja Fischer, Prof. Jens Lehmann

February 7, 2017

1 In Class

1. **Combining Models**

    a) Combining the strength of the two types of models: latent feature models are well-suited for modeling global relational patterns via newly introduced latent variables, while graph feature models are well-suited for modeling local and quasi-local graph patters.

    b) Stacking (sometimes called stacked generalization) corresponds to training a learning algorithm to combine the predictions of several other learning algorithms. While this approach has the flexibility to combine arbitrary kind of models, the models cannot cooperate, and thus need to be more complex than in combined models that are trained jointly.

2. **Closed and open world assumption**

    a) Under the closed world assumption (CWA) non-existing triples indicate false relationships. Under the open world assumption (OWA) non-existing triple are interpreted as unknown, i.e., the corresponding relationship can be either true or false.

    b) In a adjacency tensor all facts which are not existing in the corresponding KG are represented by a zero and are thus assumed to be no-existing. This corresponds to the CWA.

3. **Negative examples**

a) Known constraints which can be exploited for negative candidate generation are

- **type constraints** for predicates
(e.g., persons are only married to persons)
- **valid attribute ranges** for predicates
(e.g., the height of humans is below 3 meters)
- **functional constrains** such as mutual exclusion
(e.g., a person is born exactly in one city).

b) Irrelevant negative triples such as (*BarackObama, starredIn, StarTrek*) would not be generated by perturbation based generation since there exist no triples (*BarackObama, starredIn,* $\cdot$).

4. **Loss functions**

a) The **margin based ranking loss function**, which is given by

$$\mathcal{L}(f(x^+, \boldsymbol{\theta}), f(x^-, \boldsymbol{\theta})) = \max(0, \gamma + f(x^-, \boldsymbol{\theta}) - f(x^+, \boldsymbol{\theta})) \ .$$

encourages the score function to output higher scores for positive examples x^+ than for negative examples x^-. The loss is $\gamma + f(x^-, \boldsymbol{\theta}) - f(x^+, \boldsymbol{\theta})$ if it is positive and 0 otherwise. Since the loss is minimized there is some pressure to make $\gamma + f(x^-, \boldsymbol{\theta}) - f(x^+, \boldsymbol{\theta}) \leq 0$ and thus $\gamma + f(x^-, \boldsymbol{\theta}) \leq f(x^+, \boldsymbol{\theta})$. That means, $f(x^-, \boldsymbol{\theta})$ is encouraged to be smaller than $f(x^+, \boldsymbol{\theta})$ with a margin of γ.

b) Let the samples $x_1, \ldots, x_n$ be i.i.d. drawn from an unknown distribution which we assume to be Gaussian, i.e. we assume $x_1, \ldots, x_n \sim \mathcal{N}(\mu_0, \sigma_0^2)$ for unknown mean μ_0 and variance σ_0. Calculate the **maximum likelihood** estimate $\hat{\mu}$ of the mean. The probability density of the Gaussian distribution is given by

$$p\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Therefore, the likelihood of the parameters μ and σ given the samples $x_1, \ldots, x_n$ is given by

$$\begin{aligned}
\mathcal{L}(\mu, \sigma) &= f\left(x_1, \ldots, x_n \mid \mu, \sigma^2\right) \\
&= \prod_{i=1}^{n} f\left(x_i \mid \mu, \sigma^2\right) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right) ,
\end{aligned}$$

and the log-likelihood by

$$\log \mathcal{L}(\mu, \sigma) = -\frac{n}{2} \cdot \log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2} \ .$$

The partial derivative for μ is given by

$$\begin{aligned} \frac{\partial}{\partial \mu} \log \mathcal{L}(\mu, \sigma) &= 0 - \frac{-2\sum_{i=1}^{n}(x_i - \mu)}{2\sigma^2} \\ &= \frac{\sum_{i=1}^{n}(x_i - \mu)}{\sigma^2} \\ &= \frac{\sum_{i=1}^{n} x_i - n \cdot \mu}{\sigma^2} \end{aligned}$$

Setting this to zero and solving for mu gives

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

5. **Evaluation criteria**

a) The higher the cure the better the performance of the method. The diagonal dashed line indicates the performance for random guesses: the false positive and true positive rate are exactly the same.

b) The right candidate e_2 has the 3rd largest score. Thus, the reciprocal rank of the model is $\frac{1}{3}$.