Computer Science Department, University of Bonn

Knowledge Graph Analysis

Exercise Sheet 10

Dr. Asja Fischer, Prof. Jens Lehmann

February 7, 2017

1 In Class

1. **Combining Models**

   a) What is a reason for combining latent feature and graph feature models?

   b) What is stacking?

2. **Closed and open world assumption**

   a) What is the closed, what is the open world assumption?

   b) Do tensor factorization models rely on the closed or open world assumption?

3. **Negative examples**

   a) What kind of known constraints can be exploited for negative example generation?

   b) Explain why defining the set of negative examples as

$$\begin{aligned}\mathcal{D}^- = &\{(e_l, r_k, e_j) | e_l \neq e_i \wedge (e_i, r_k, e_j) \in \mathcal{D}^+\} \\ &\cup \{(e_i, r_k, e_l) | e_l \neq e_j \wedge (e_i, r_k, e_j) \in \mathcal{D}^+\}\end{aligned}$$

   leads to focus on more plausible negative examples.

4. **Loss functions**

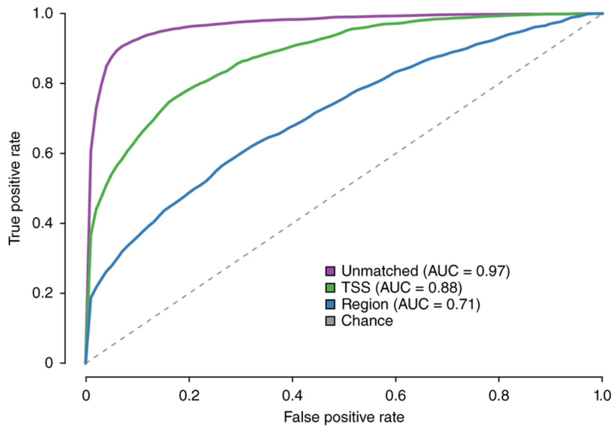
a) Explain the **margin based ranking loss function**, which is given by

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

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.

5. **Evaluation criteria**

a) Look at the **receiver operating characteristic (ROC)** curves in the following figure[1]. Which method is the best? And what is indicated by the dashed line?



b) Assume the entities e_1 and e_2 correspond to the same object (i.e. the entitiy is duplicated). Presenting e_1 to en entity resolution model it outputs the following scores for a list of candidates

entities	score
e_2	2.34
e_3	3.47
e_4	2.1
e_5	1.7
e_6	1.83
e_7	2.89
e_8	2.01
e_9	1.69

What is the **reciprocal rank** of the model?

[1] The figure was taken from *Ritchie et al.,Functional annotation of noncoding sequence variants, Nature Methods 11, 294-296 (2014)*