Master Thesis

Kavita Chopra

presented on 21.12.2016

Superviser: Denis Lukovnikov

Reformulation of Topic:

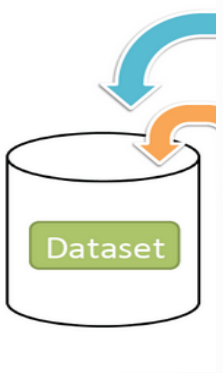
Evaluation of characteristics of Knowledge Base Completion Models

(context: question answering using web data)

Quick recap of status quo

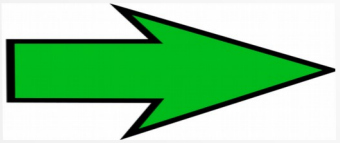
- Representation Learning for Knowledge Graph Embedding
  - TransE
  - Rescal

own implementation using Tensorflow



FB15K
(600,000 triples)

  - Data set: Freebase-subset
  - Widest use for:
    - inference of new facts from existing facts in KB
    - Cleaning KB: detection of false facts



Knowledge Graph Completion

Evaluation of model perfomance

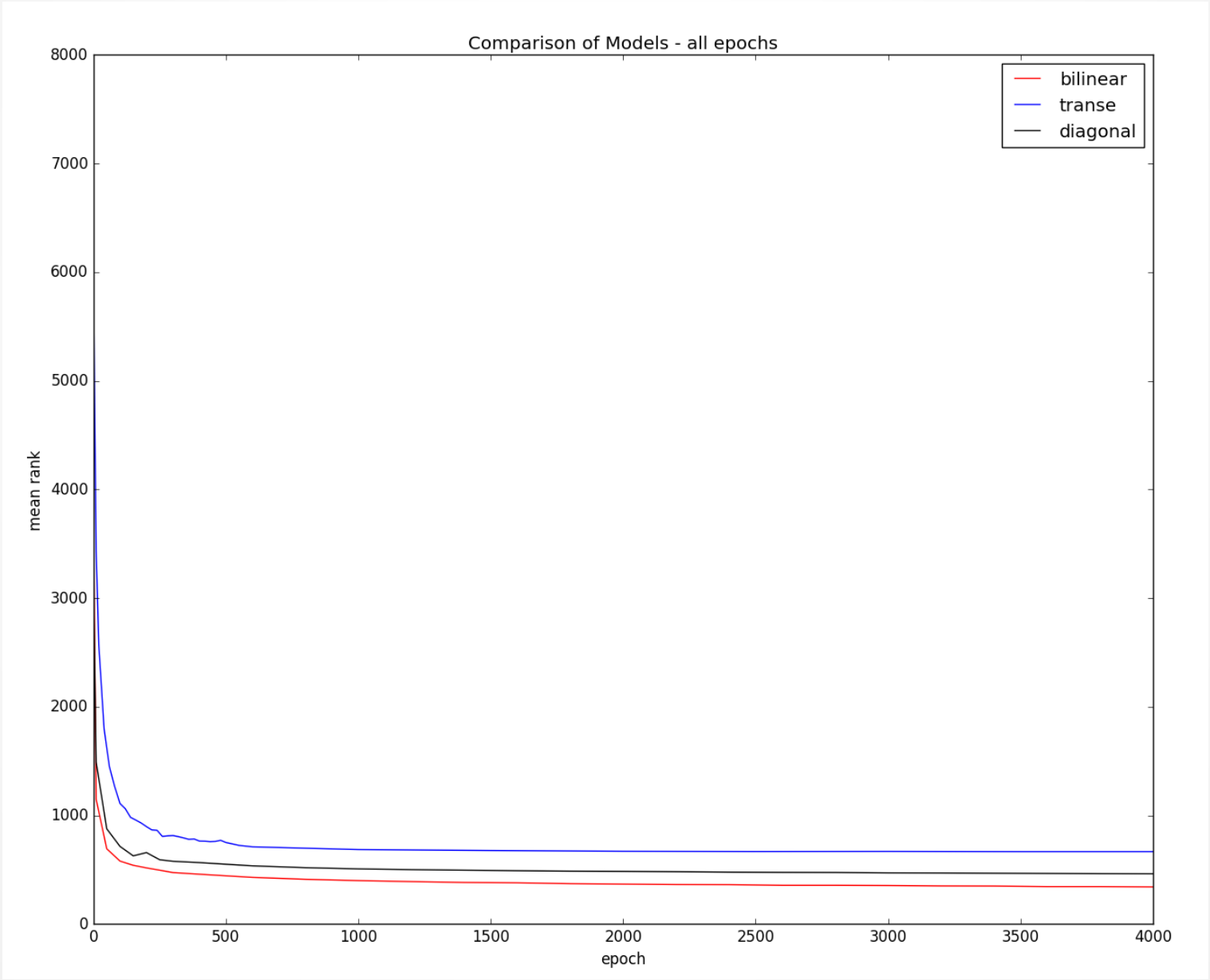
Two tasks:

- Link Prediction & classification task

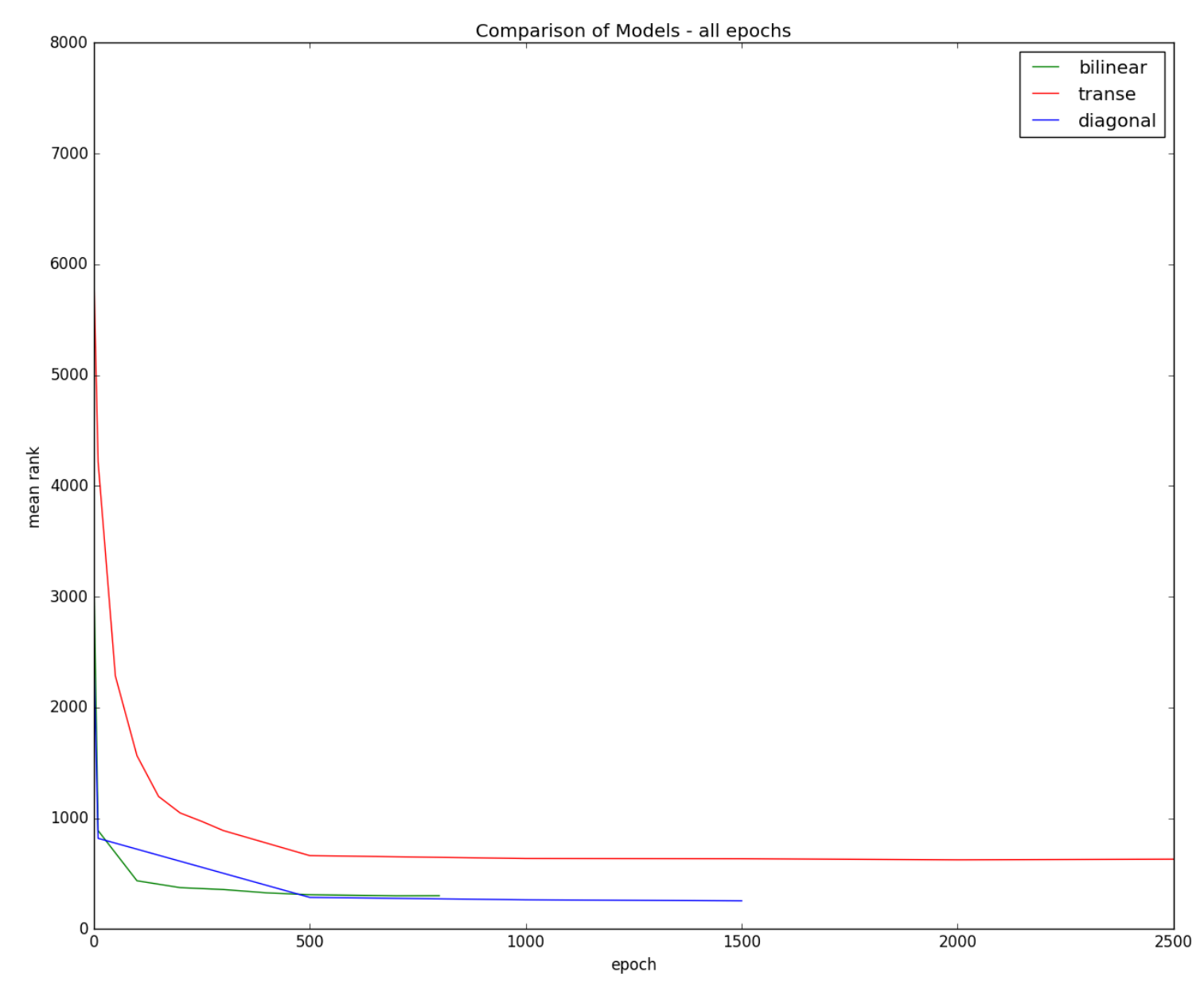
Link prediction:

- Given head and label, predict tail
- Given label and tail, predict head

TransE, Bilinear, Diagonal (dim=20)



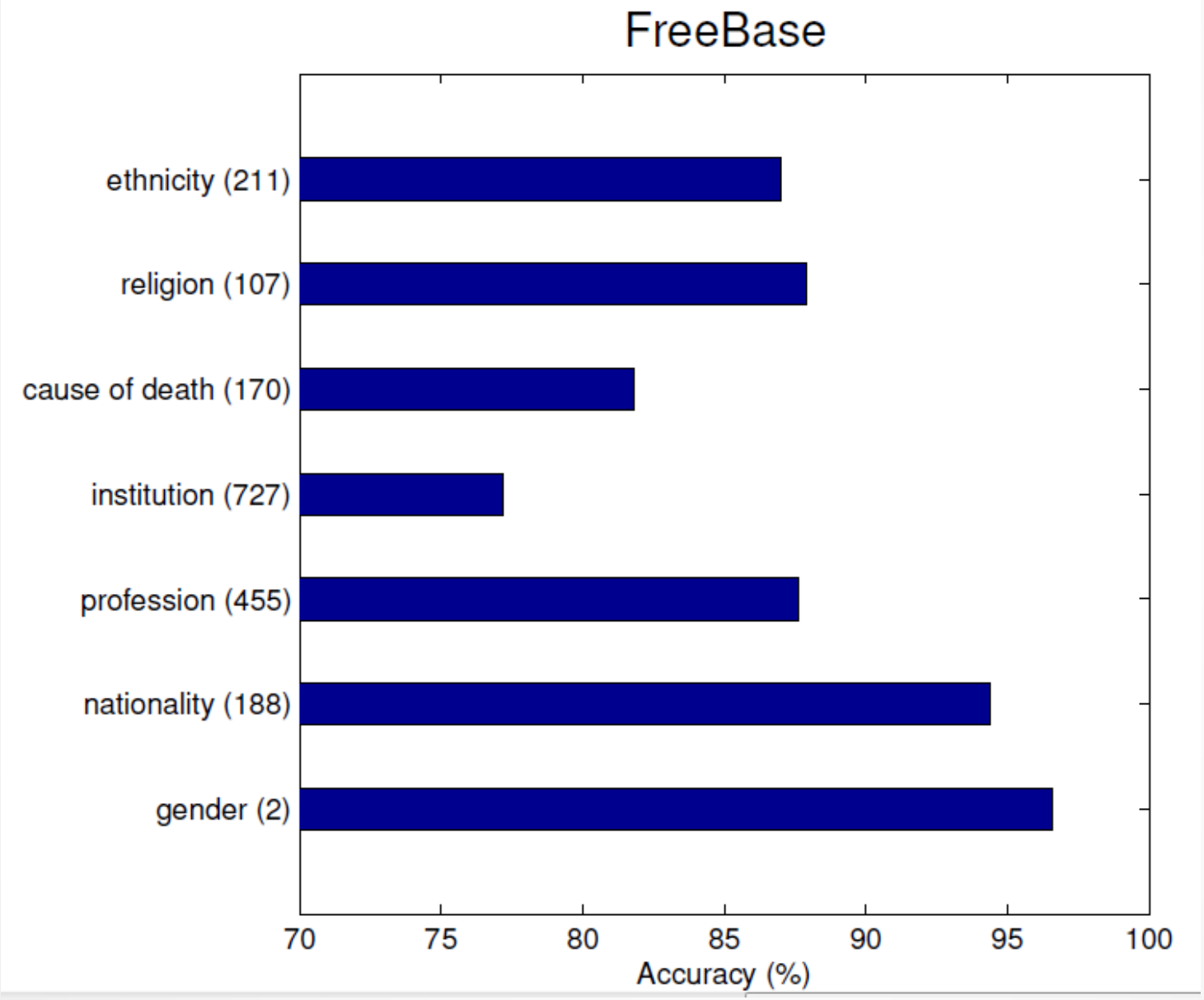
TransE, Bilinear, Diagonal (dim=100)



Evaluation: Classification

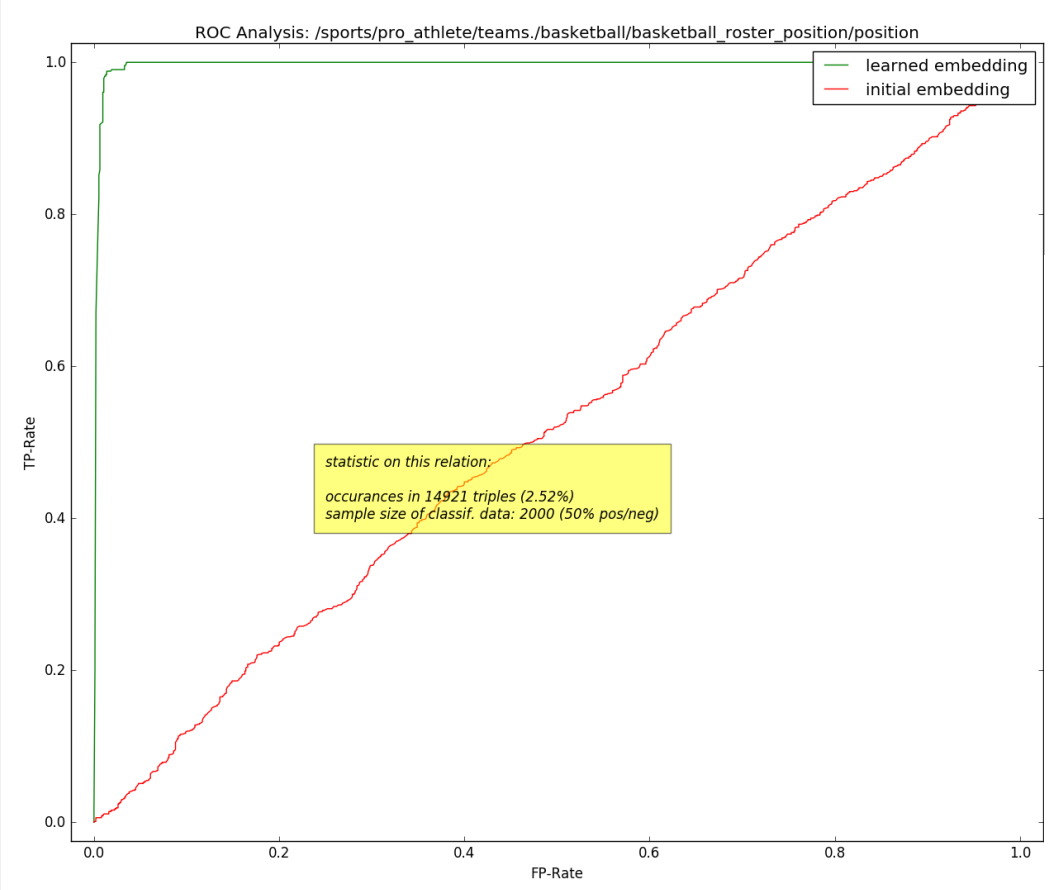
- For every relation sample a set of triples (positive!)
- Create a negative set by replacing head or tail with random, but representative entity, e.g.
  - Positive: Obama /presidentOf US
  - Negative: Obama /presidentOf Germany
- Compute scores and using a threshold t* derive t the confusion matrix and report accuracy

Evaluation: Classification

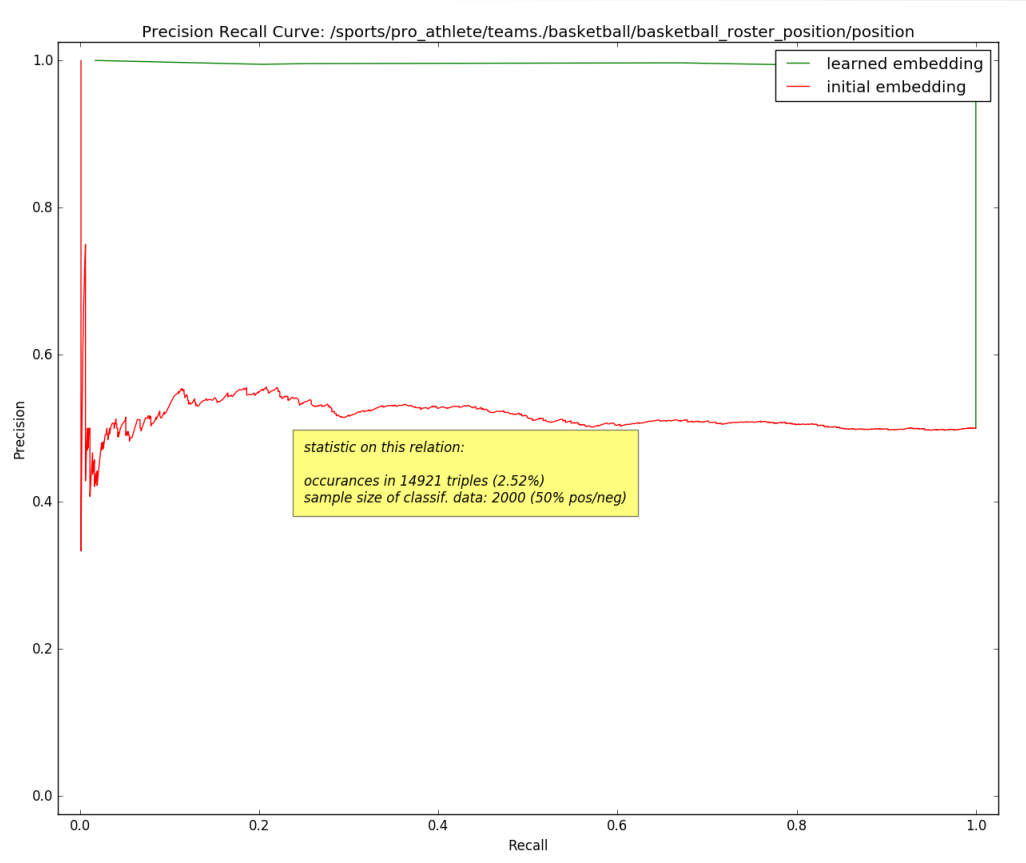


Procedure in literature

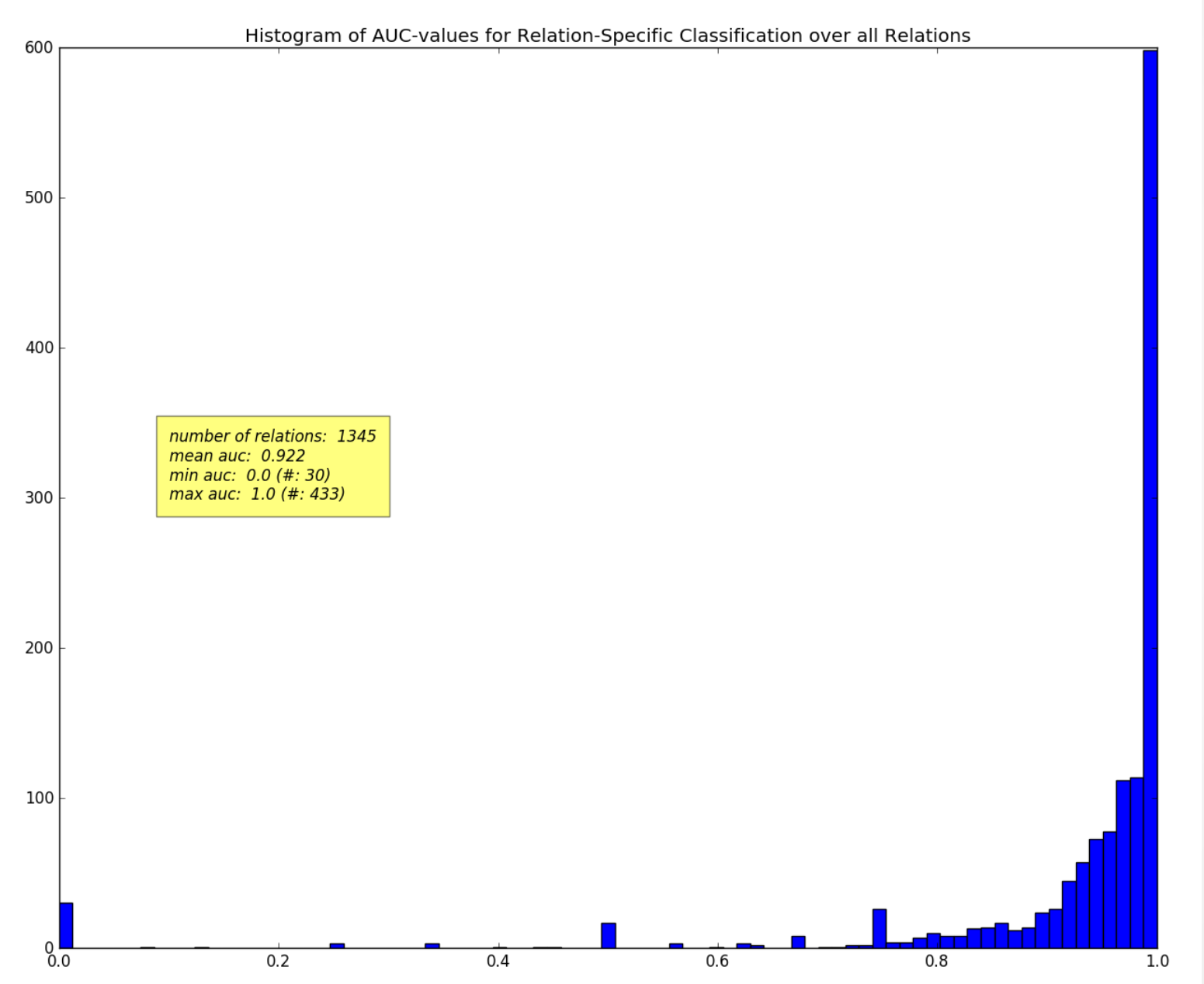
Classification - my approach:



Find best threshold via ROC-analysis



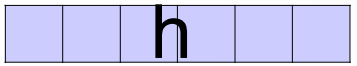
Classification - my approach:



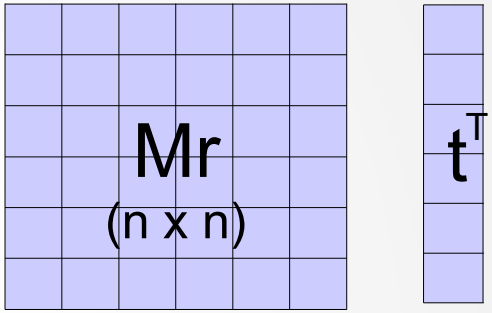
What‘s next?

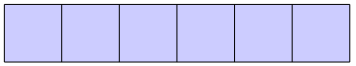
- Bilinear: h * Mr * t^T
- Bilinear is powerful but many parameters...
- diagonal is not bad either…
- What about a middle way?
- Why not decompose Mr?
  - Mr = A * B^T

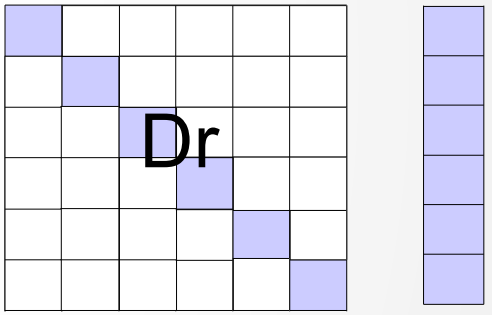
    (A, B are n x a where 1 <= a <= n/2)

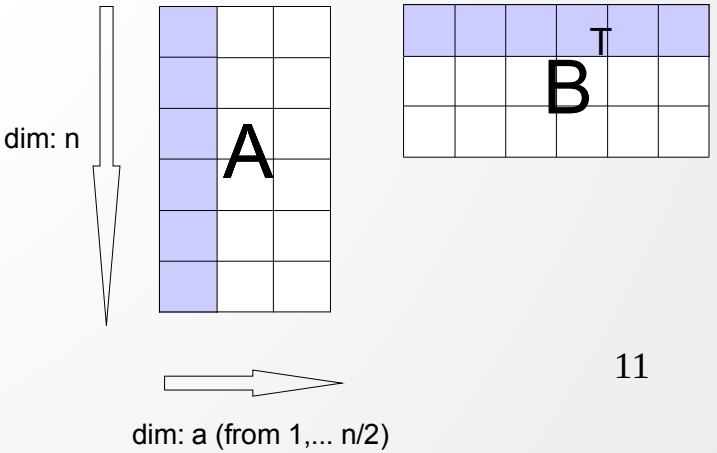


n is a low latent dimension










THANK
YOU