Master Thesis

Kavita Chopra

presented on 19.10.2016

Superviser: Denis Lukovnikov

Topic:

Prediction of the next entity/relation in a sequence of triples derived from Knowledge Graphs using Recurrent Neural Networks

(context: question answering using web data)

Knowledge Base Completion

- Representation Learning
  - TransE
  - Rescal

TransE Algo

Algorithm 1 Learning TransE

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

initialize $\ell \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

  $\ell \leftarrow \ell / \|\ell\|$ for each $\ell \in L$

  $\mathbf{e} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

  $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

  $S_{batch} \leftarrow$ sample(S, b) // sample a minibatch of size b

  $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

  for $(h, \ell, t) \in S_{batch}$ **do**

    $(h', \ell, t') \leftarrow$ sample$(S'_{(h,\ell,t)})$ // sample a corrupted triplet

    $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

  end for

  Update embeddings w.r.t. $\sum_{\left((h,\ell,t),(h',\ell,t')\right) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$

end loop

$$\mathcal{L} = \sum_{(h,\ell,t) \in S} \sum_{(h',\ell,t') \in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$$

TransE

$$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} [\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h}'+\boldsymbol{\ell},\boldsymbol{t}')]_+$$

- **Hyperparameters:**
  - Margin: {1,2,10}
  - Learning rate: {0.001, 0.01, 0.1}
  - Dimension k: {20,50,100}
  - Distance measure: L1 or L2 norm
  - Batch size: 100
  - Max-epoch bound: 1000, early stopping
  - For corruption of triples: collision check

- **Optimal**
  - **Margin: 2**
  - **Learning rate: 0.01**
  - **Dimension k: 20**
  - **Distance measure: L1**

TransE

- TransE (Bordes, Antoine, et al. "Translating embeddings for modeling multi-relational data." Advances in Neural Information Processing Systems. 2013.)
  - Distance based model which models relations between entities as translations in the embedding space

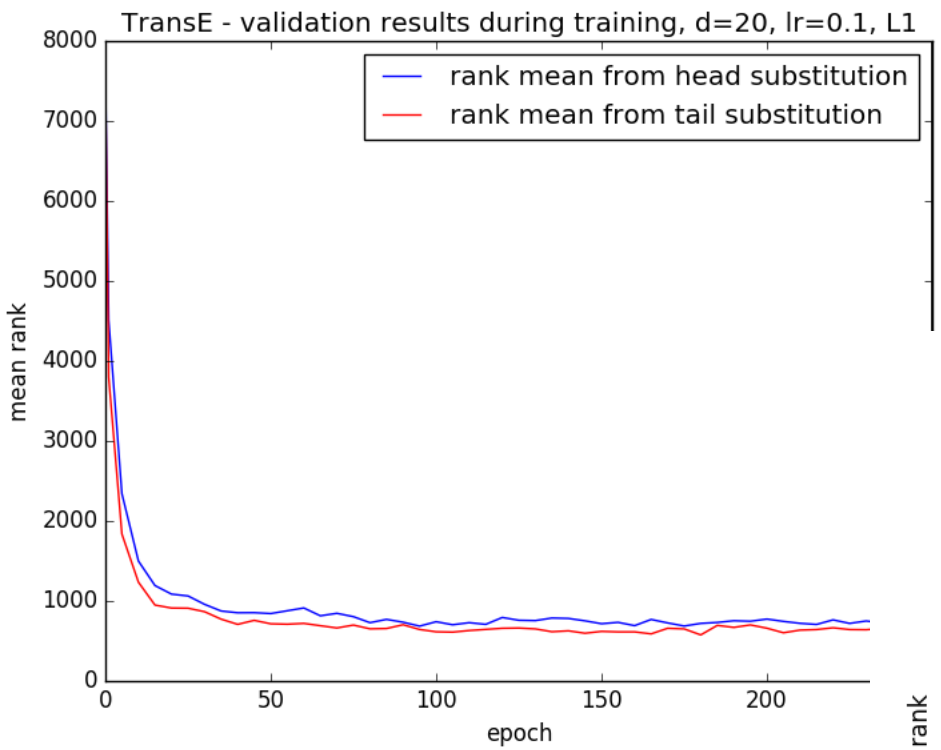
TransE Training Data

- Freebase: > 1.2 billion triples, > 80 million entities
- FB15k:
  - 600,000 triples split into
    - ~500,000 training triples
    - ~50,000 test tripes for evaluation
    - ~50,000 validation triples data
  - ~15,000 entities, ~1,400 relations

TransE Validation and Evaluation Protocol

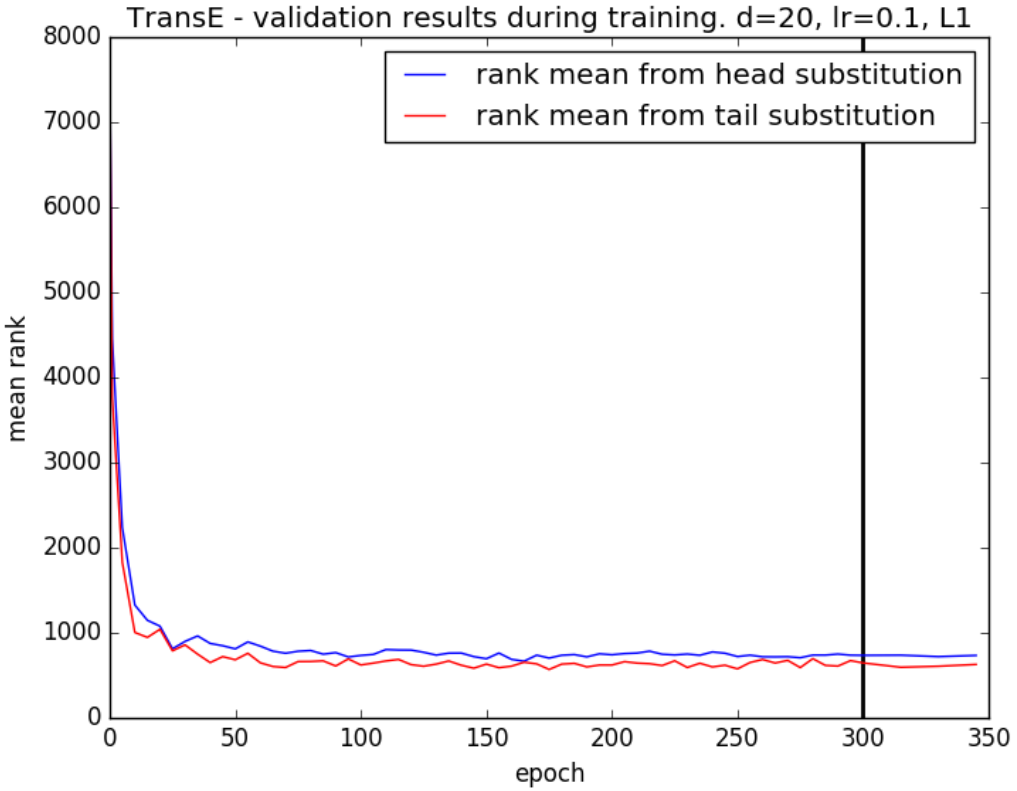
- Validation: with valid-set (50,000 tr.)
  - for all valid-set triples, fix head and label, iterate over all possible entites to substitute the correct tail and compute score:
    - score_ji = d(head_fixed_j+label_fixed_j – tail_ji)
    - (j: index from list of valid-set triples, i: index from list of entities)
  - Repeat same for fixed tail and label
  - Sort resulting lists of scores and report the mean rank of the score of the correct triple
- Evaluation: with test-set (50,000 tr.)
  - Repeat the same for all test-triples to get the mean ranks + check for every 10 ranked triples whether in set of all triples (~600,000)
  - Report the proportion of the hits@10

TransE: dim=20, LR= 0.1, L1



Rank mean from head replacement:
717 (out of 14,951)
Hits@10: 21.696 %

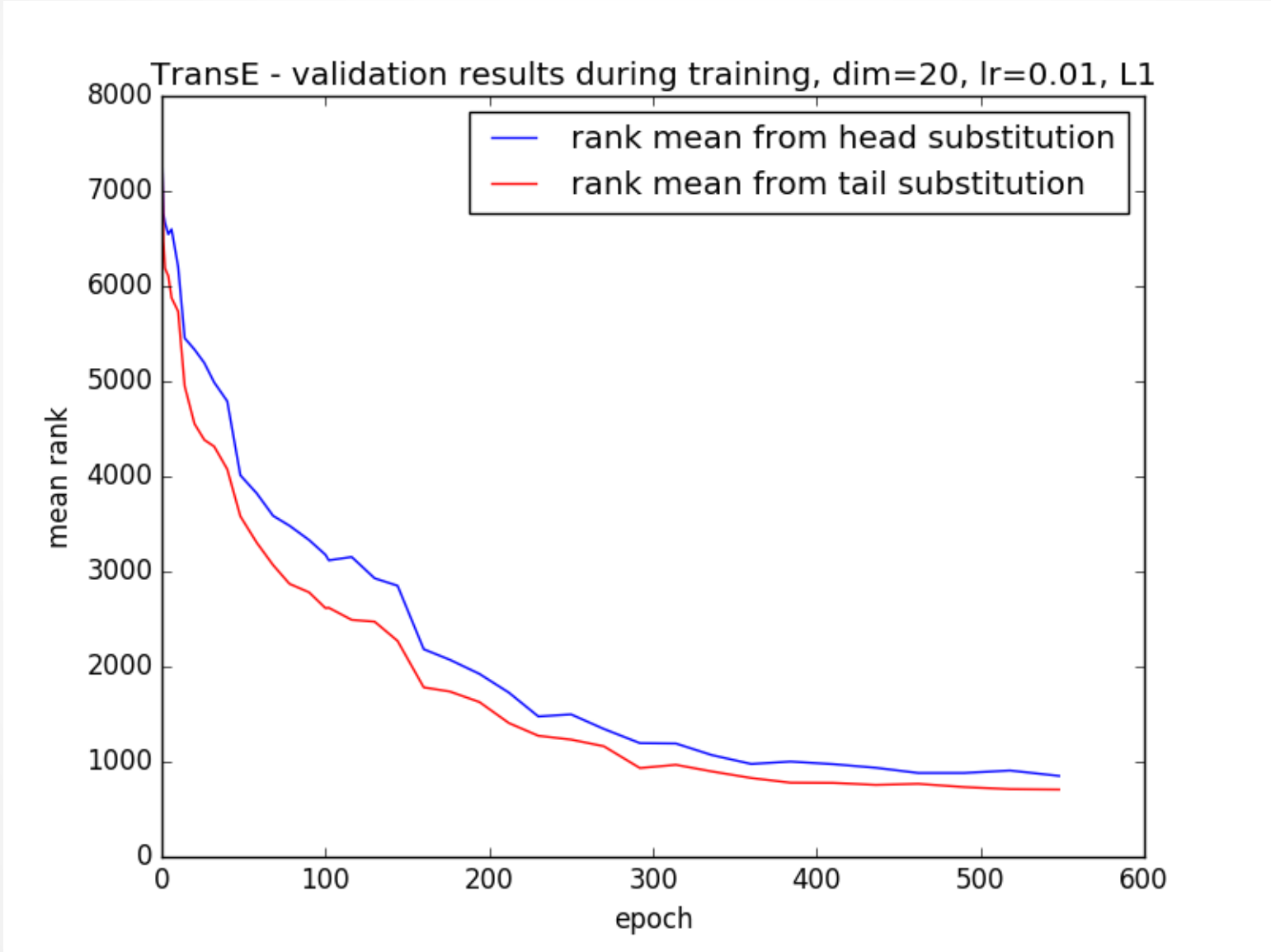
Rank mean from head replacement: 612
(out of 14,951)
Hits@10: 17.72 %



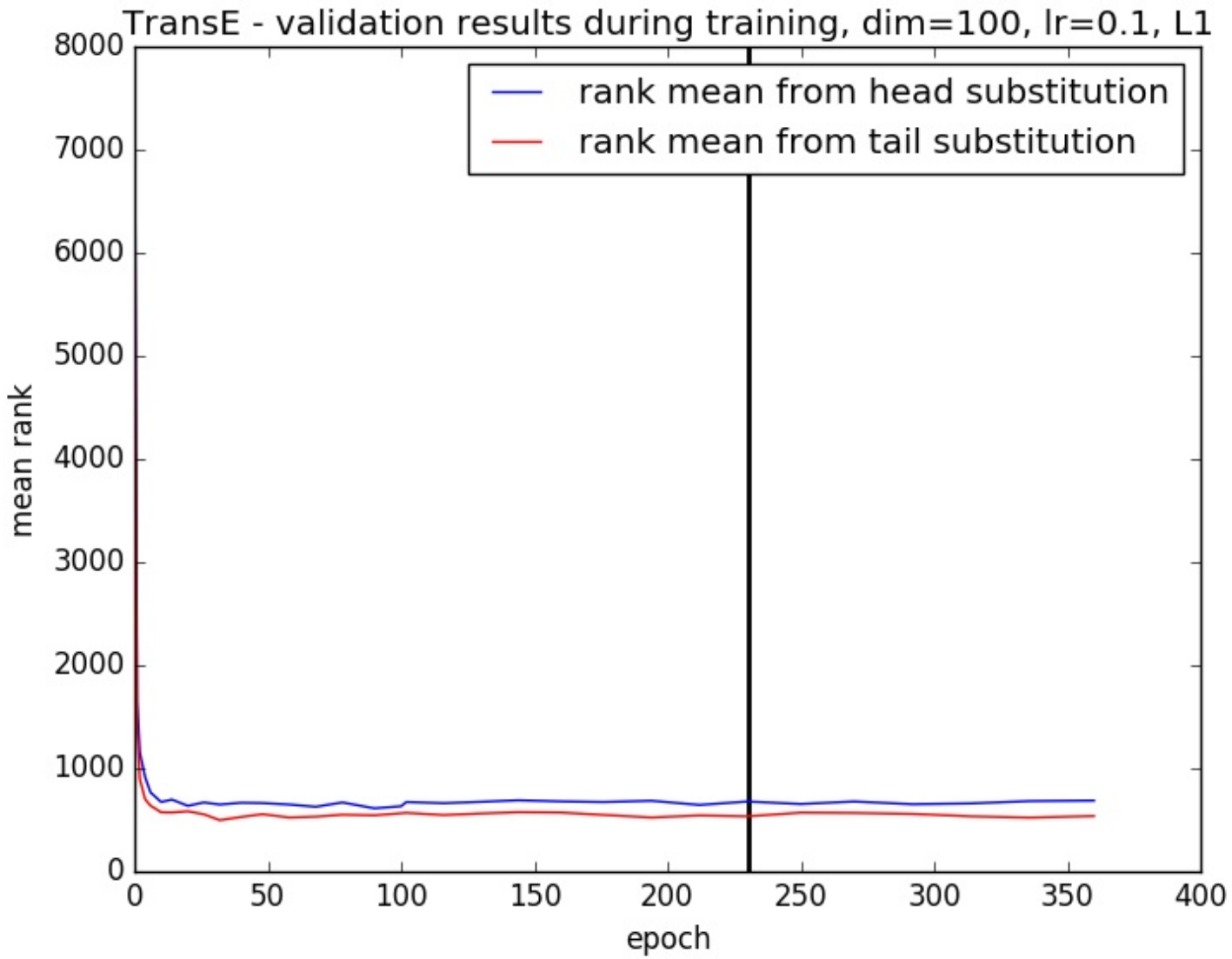
Left:
model trained with 1 negative triple per positive

Right:
model trained with 5 negative triples per positive, alternated after every epoch

TransE: dim=20, LR= 0.01, L1



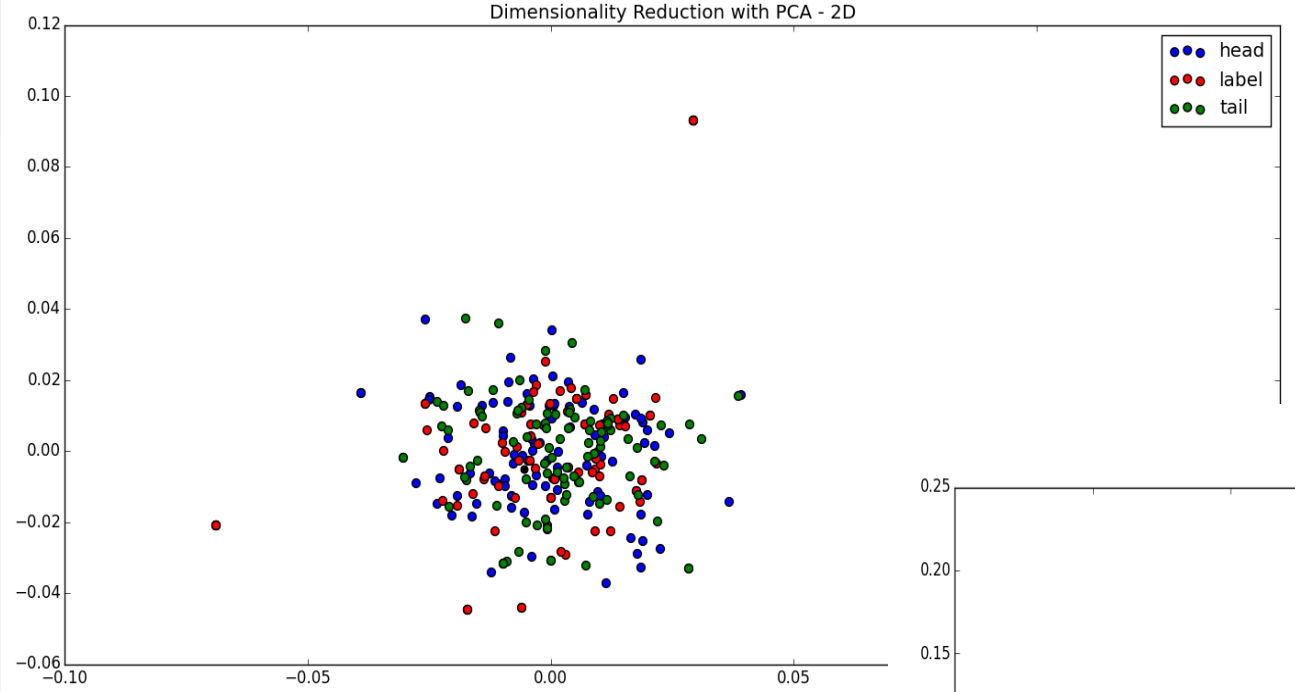
TransE: dim=100, LR= 0.1, L1

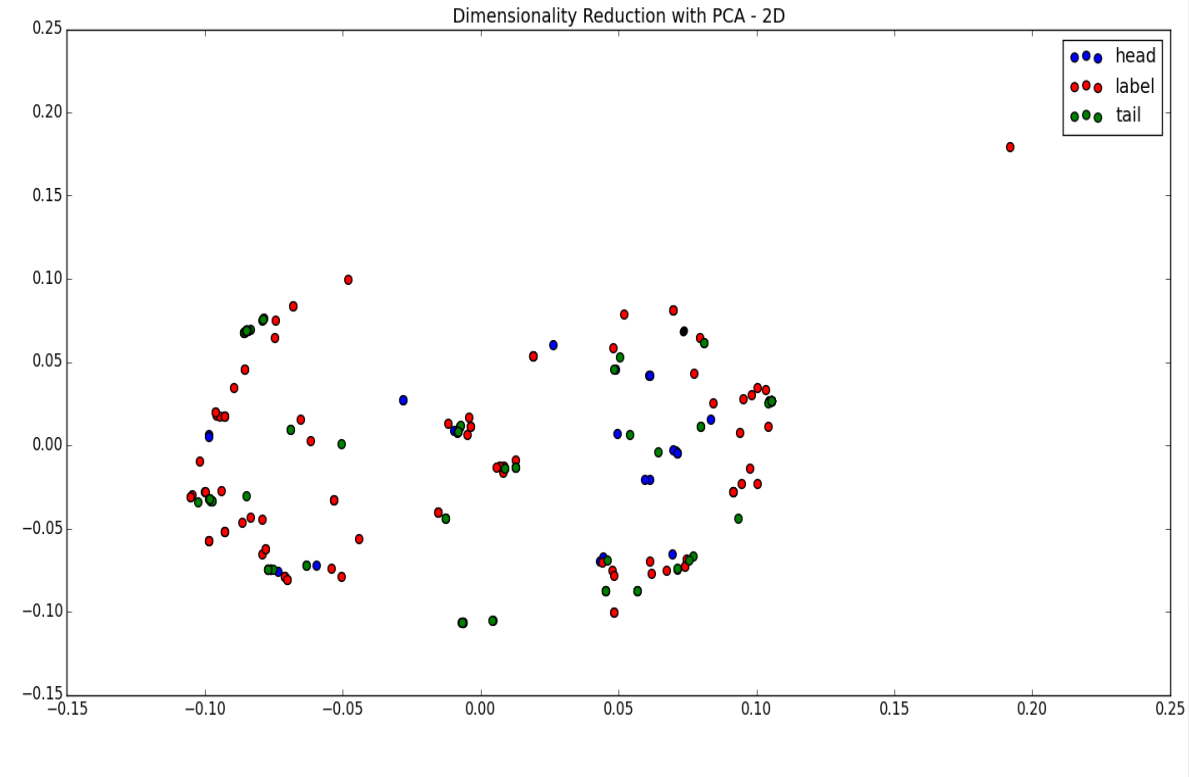


Rank mean from head replacement:
660 (out of 14,951)
Hits@10: 23.802 %

Rank mean from head replacement:
572 (out of 14,951)
Hits@10: 20.486 %

PCA with TransE embeddings

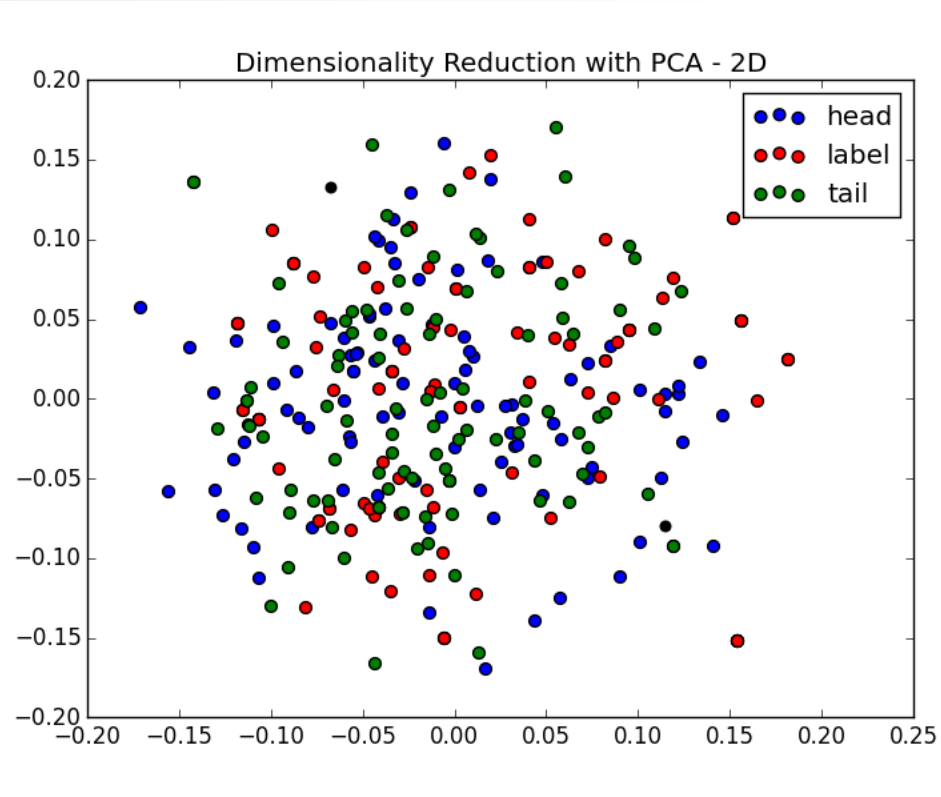


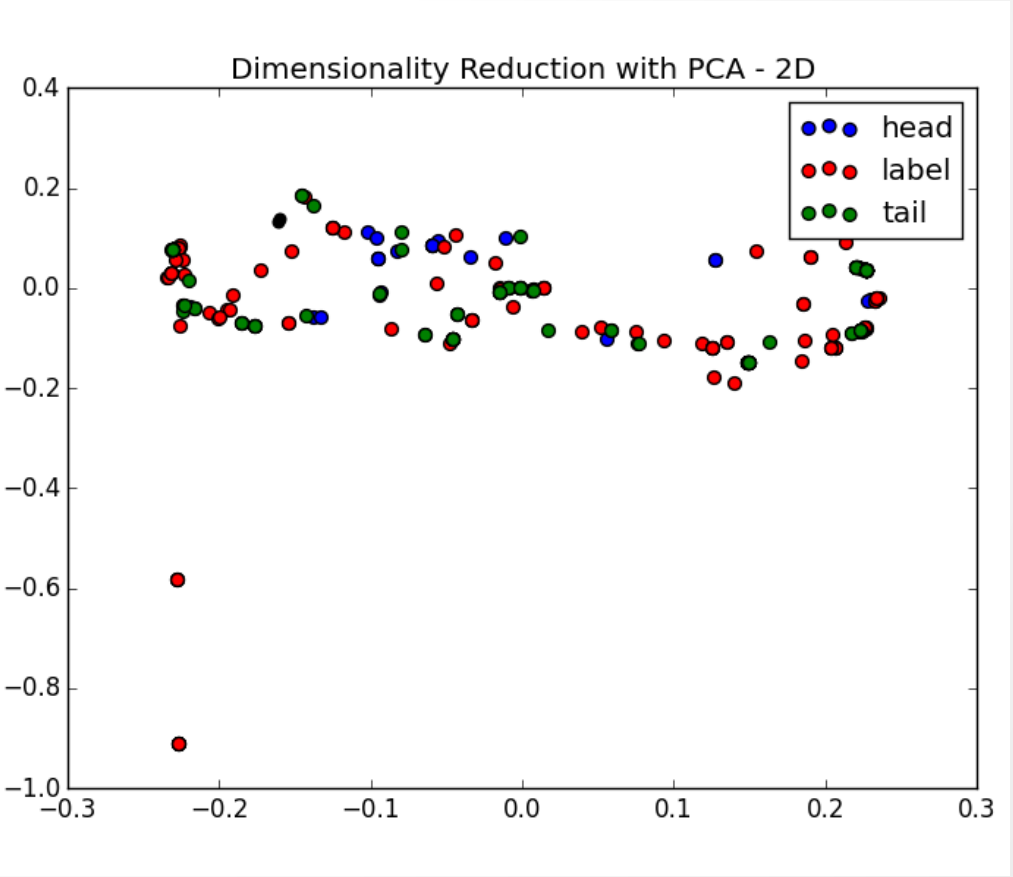


Left:
Initial representation for dim = 100

Right:
Learned emebdding for dim= 100

PCA with TransE embeddings





Left:
Initial representation for dim = 20

Right:
Learned emebdding for dim = 20

What‘s next?

- Will try to improve runtime of implementation to run more experiments with different configurations
- Implementation of Bilinear/ RESCAL model


THANK
YOU