Paraphrase Embeddings

Yiyi Chen
Profilierungsmodul II
Dozent: Dr. Benjamin Roth

Task

Learning embedding vectors for syntactic **dependency paths** between entities.

arg0:m.0_00 dobj^-__accuse__nsubj arg1:m.07q13

⇒ m.0_00#m.07q13 dobj^-__accuse__nsubj

Dataset

Data (paths occur at least once in **test_simplified.csv**, in total **1759234** instances, **num of path**: 1717):

arg0:m.0_00 dobj^-___accuse___nsubj arg1:m.07q13

⇒ **m.0_00#m.07q13 dobj^-___accuse___nsubj**

Processed data : (entity pairs occur at least three times, negative factors 5)

Entity pairs	Paths	**samples**	**With negative samples**
66835	1714	410370	2051850

Train Dataset & Dev Dataset

dobj^-___join___nsubj m.0_00#m.01gf5z m.05cgv#m.0d05l6

dobj^-___join___nsubj m.0_00#m.01gf5z m.0dtj5#m.045c7b

dobj^-___join___nsubj m.0_00#m.01gf5z m.01mrvg#m.044qx

dobj^-___join___nsubj m.0_00#m.01gf5z m.0g0c_#m.0gwlg

dobj^-___join___nsubj m.0_00#m.01gf5z m.03qpbs#m.01xyt7

Train samples: 2031332 (99 %)

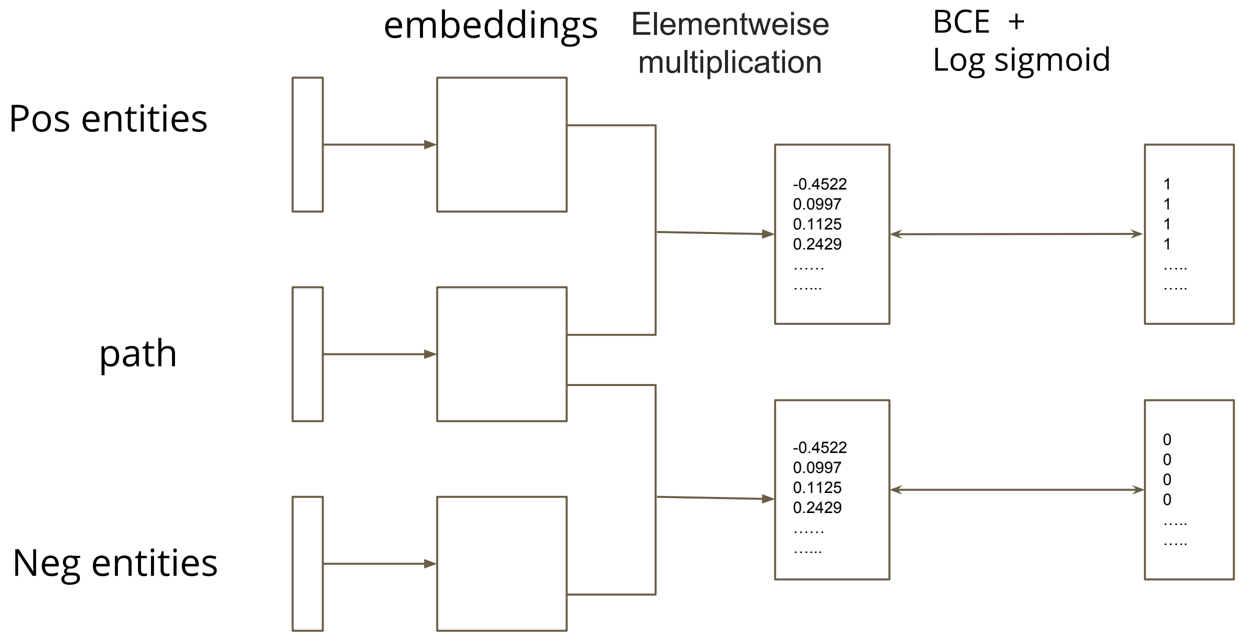
Dev samples: 20518 (1%)

Word2Vec

Objective:

$$sigmoid\{pos\_ents^\top path\} \sim 1$$

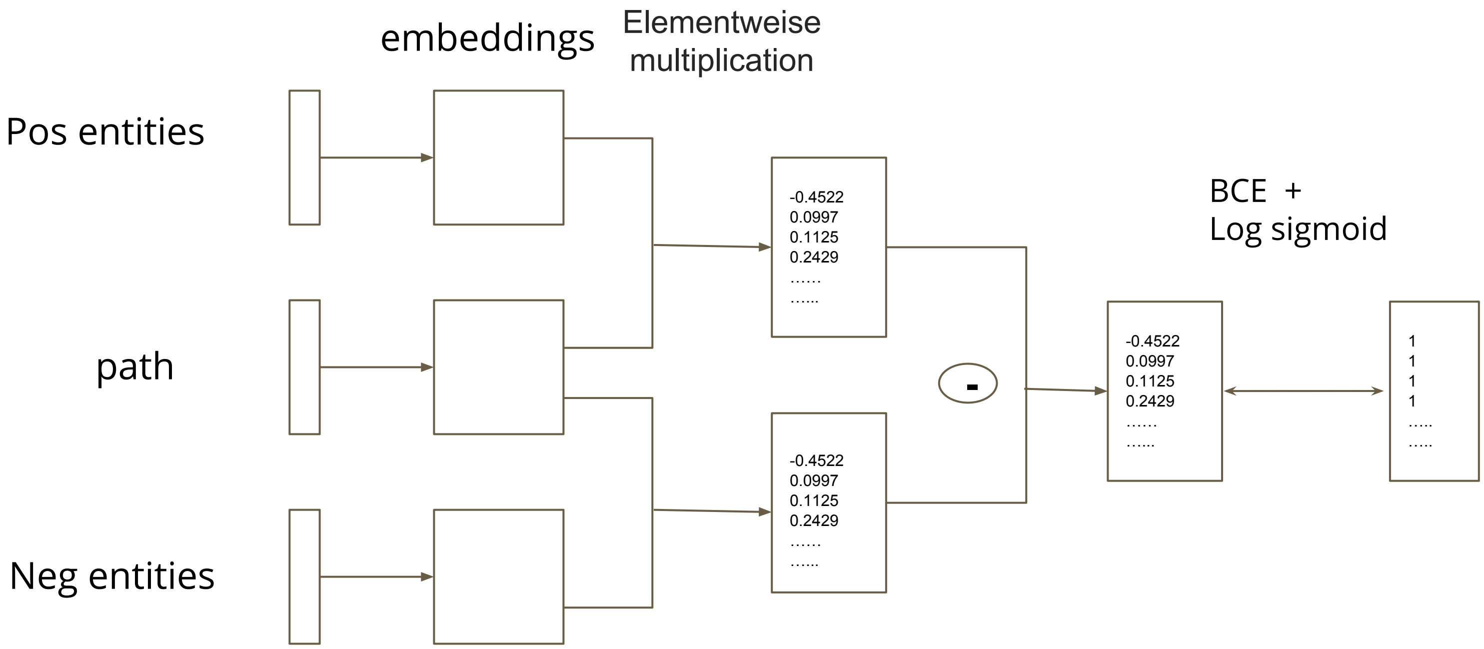
$$sigmoid\{neg\_ents^\top path\} \sim 0$$



Universal Schema

Objective:

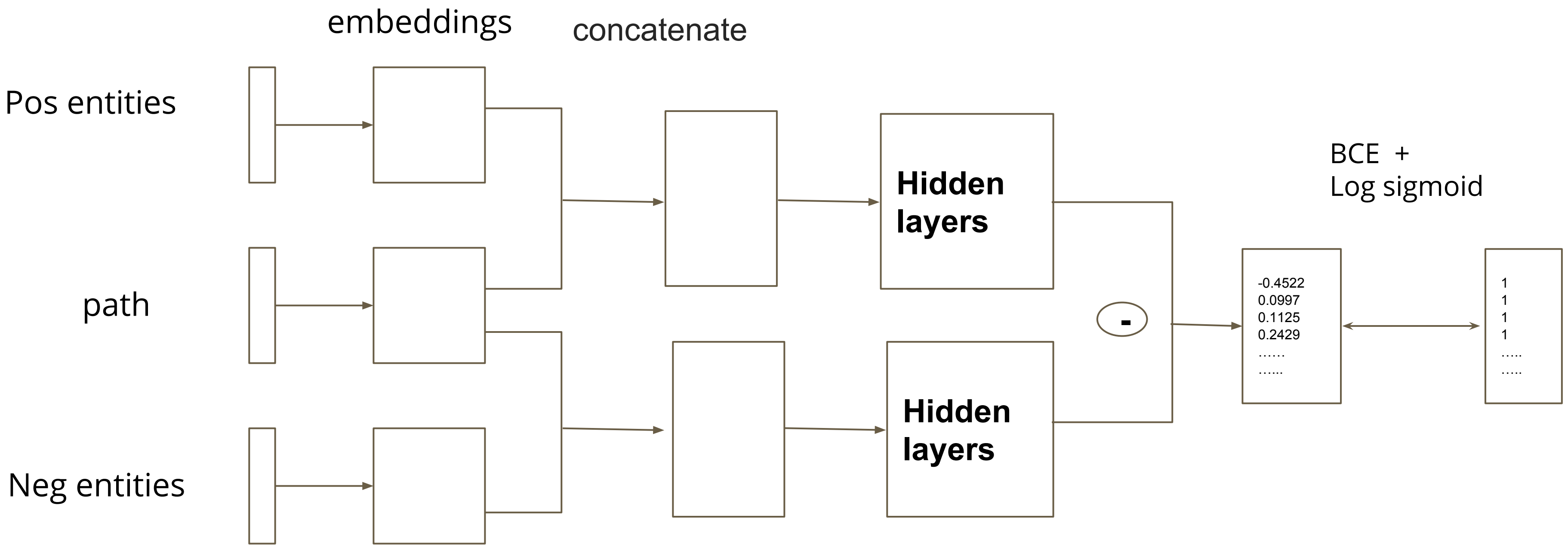
$$sigmoid\{pos\_ents^\top path - neg\_ents^\top path\} \sim 1$$



Sequential Model

Objective:

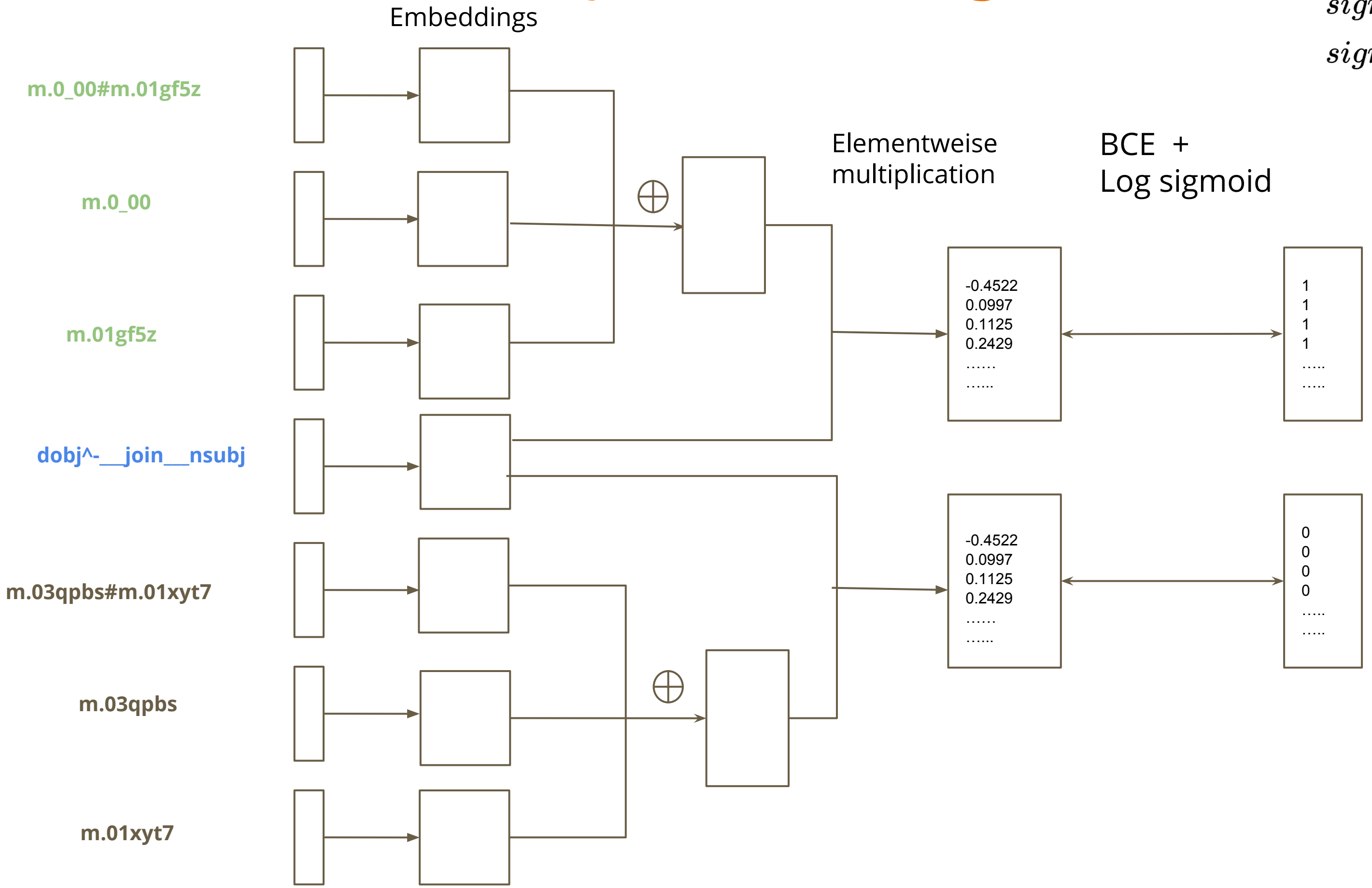
$$sigmoid\{pos\_ents^\top path - neg\_ents^\top path\} \sim 1$$



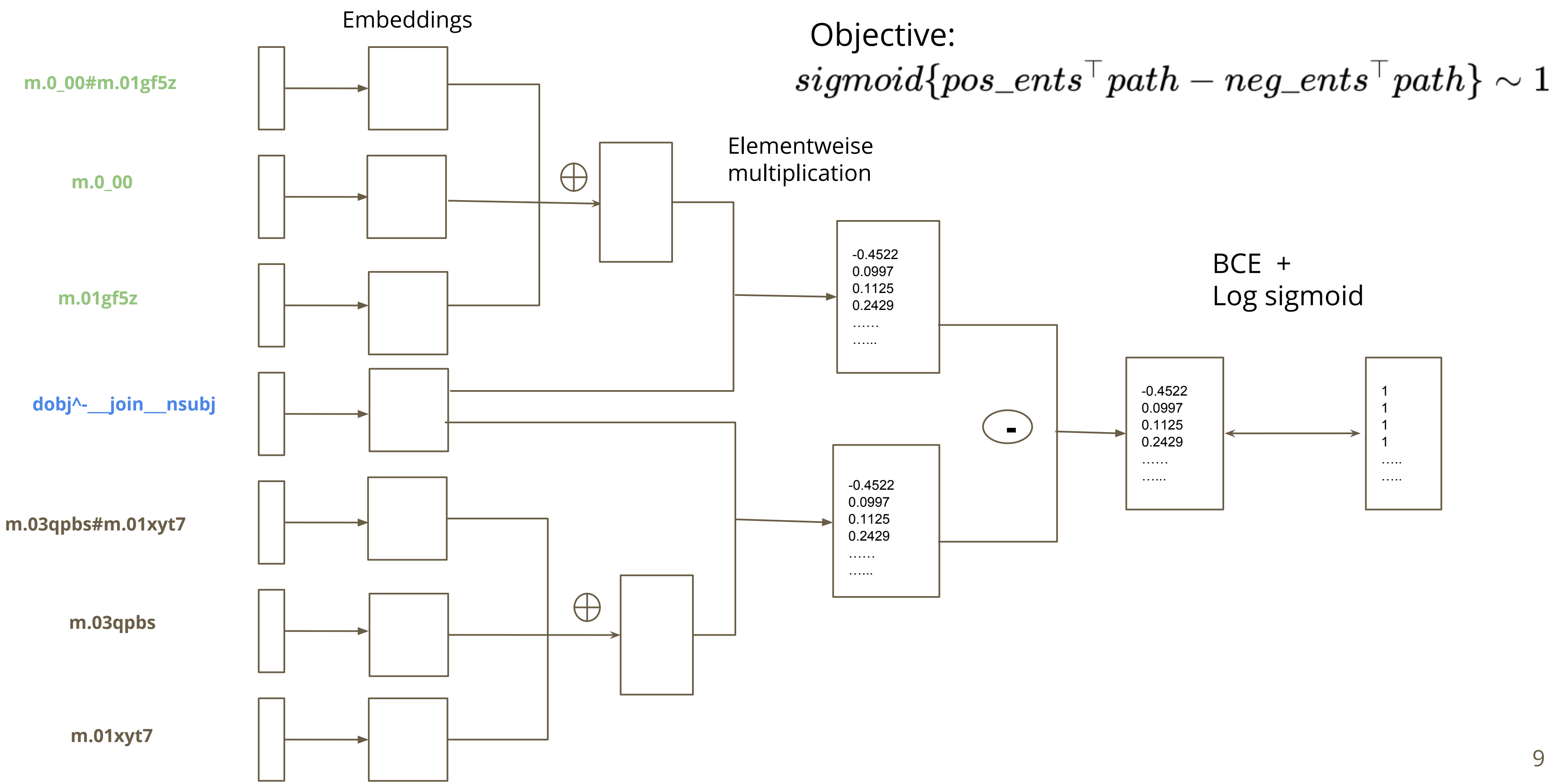
Word2Vec + entity embeddings

Objective:

$$sigmoid\{pos\_ents^\top path\} \sim 1$$
$$sigmoid\{neg\_ents^\top path\} \sim 0$$



Universal Schema + entity Embeddings



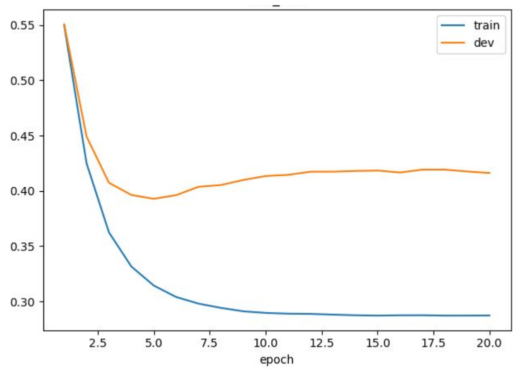
Hyperparameters

Batch-size: 4096

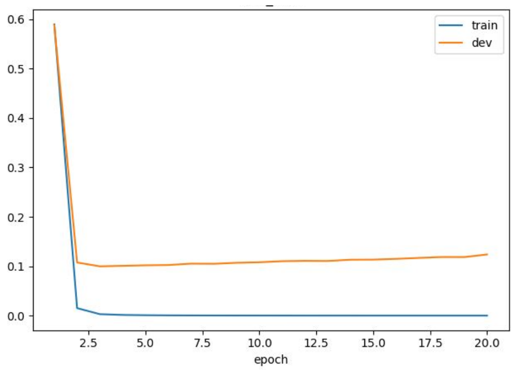
Num of epochs : 20, 30, 40, 80

Optimizers: sgd, adadelta, adam (amsgrad)

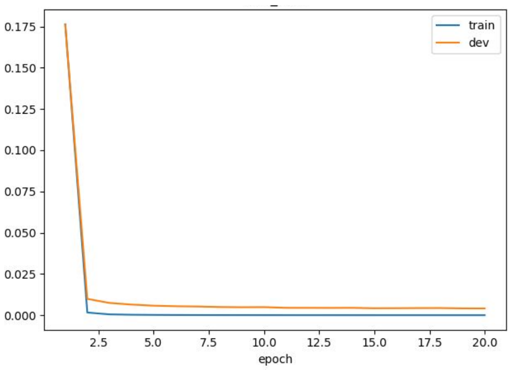
Learning rate: 0.001, 0.01, 0.1



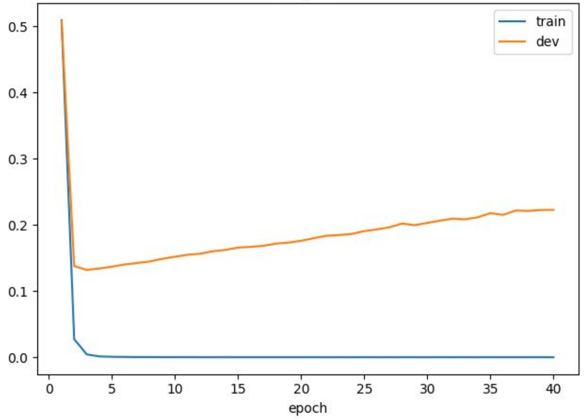
Sequential Model
(1 lstm, 1 linear layer, 20 epochs, adam 0.01)



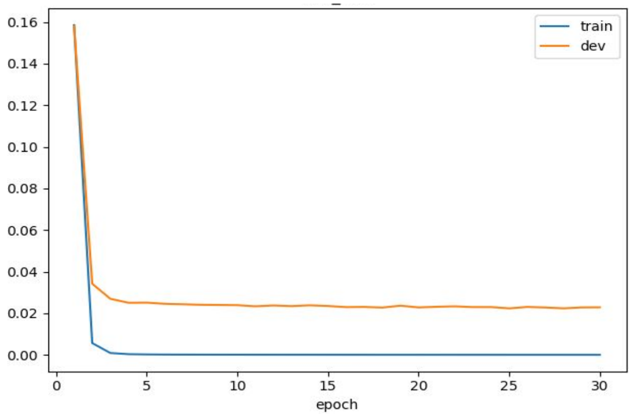
Word2vec (20 epochs, adam 0.01)



Universal Schema (20 epochs, adam 0.01)



Word2Vec +entity_embeddings
(40 epochs, adam 0.01)



Universal Schema +entity_embeddings
(30 epochs, adam 0.01 amsgrad)

Result

	Mean Precision	Mean Recall	Mean F1 Score
Baseline	0.471	0.712	0.565
Word2Vec	**0.566**	0.720	0.633
Word2Vec +entity_embeddings (40 epochs, adam 0.01)	0.518	0.777	0.621
Universal Schema	0.549	0.787	**0.646**
Universal Schema +entity_embeddings (30 epochs, adam 0.01 amsgrad)	0.507	0.796	0.618
Sequential Model	0.418	**0.809**	0.551

Remarks

- Data samples structured as path pos_entity_pair is much less memory consuming than path pos_entity_pair True and path neg_entity_pair False
- To report the dev_loss, detach the tensor and only record the number loss_dict[“dev”].append(dev_loss.detach().item()) to avoid out of memory error
- To concatenate the embeddings of entity1 and entity2 in order to get the embeddings of entity1#entity2 does not work well for training the models, better to use embeddings of entity1#entity2 directly or embeddings of entity1#entity2 plus embeddings entity1 and embeddings of entity2.
- The model with hidden layers does not work better than word2vec and universal schema, maybe due to the very short sequence length 2 (path pos_entity_pair or path neg_entity_pair as one sequence).

References

Sebastian Riedel, Limin Yao, Andrew McCallum, Benjamin M.Marlin, *Relation Extraction with Matrix Factorization and Universal Schemas*, Proceedings of NAACL-HLT 2013