

Hypernym discovery : CRIM experiments(6)

CRIM 구현 및 실험

Geonil Yun,

rjsdlf45@gmail.com

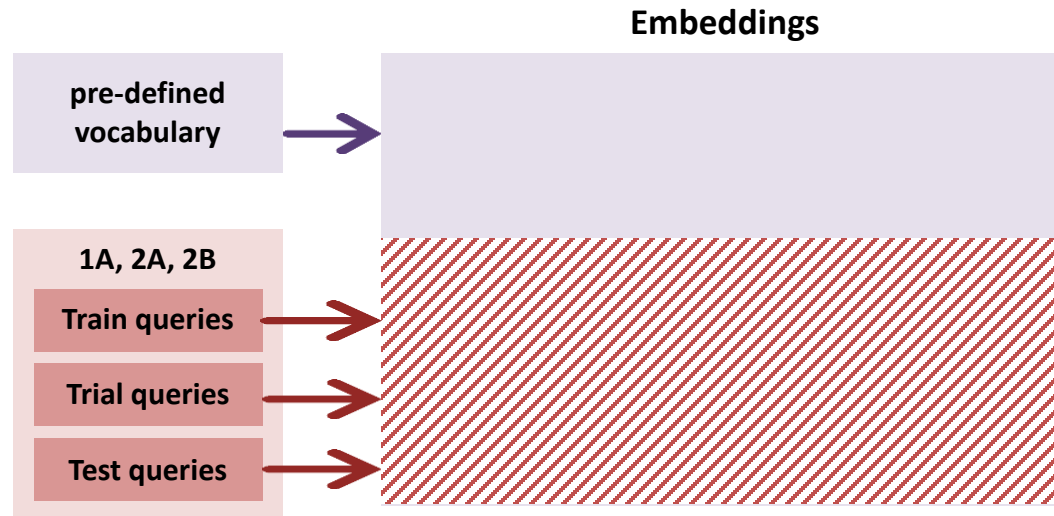
2021. 11. 02.



Experiment setting

- Embeddings : Add queries

- Embeddings vocab

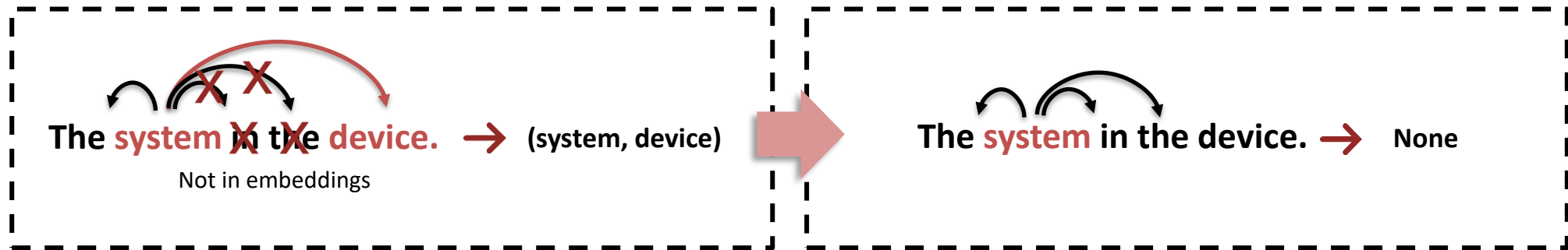


- Previous embeddings are constructed of **pre-defined vocabulary**.
- Addition **training, trial, test queries** of 1A, 2A, 2B dataset to existing embeddings.

Experiment setting

- Word Embeddings

- Skip-gram model structure



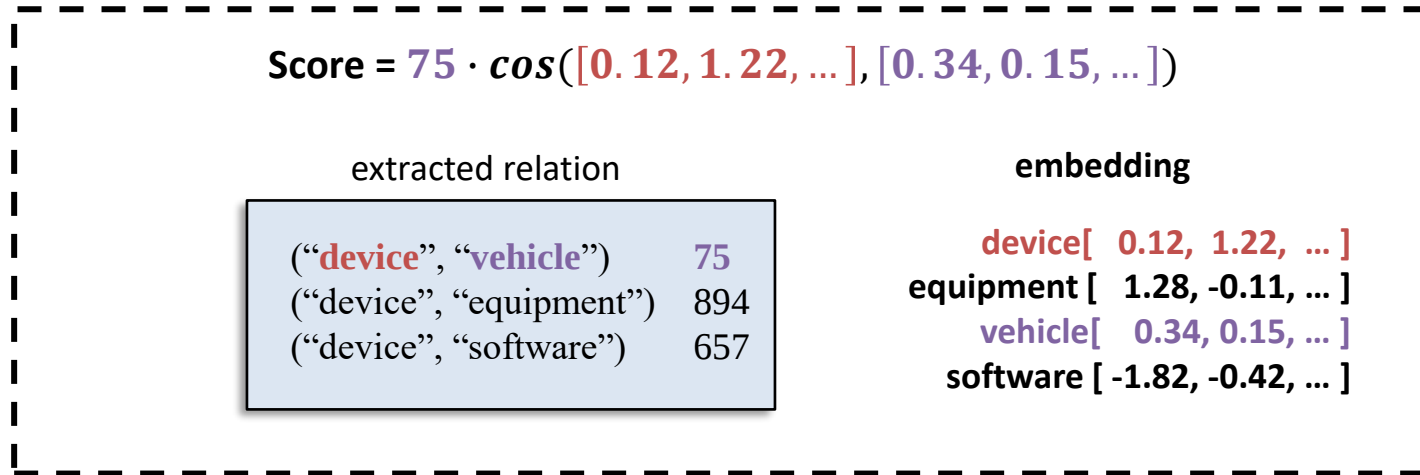
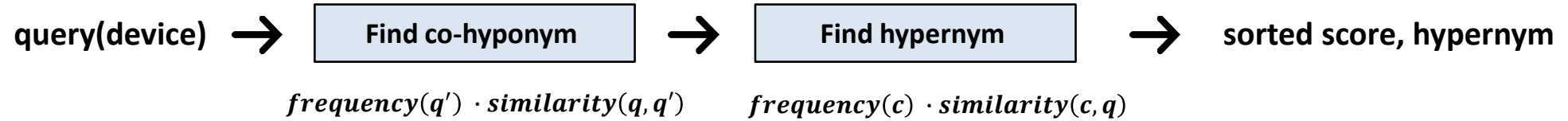
hyperparameter	setting
embedding dimension	100
negative samples	15
batch size	300,000
learning rate	0.0001
epoch	100
window size	2, 3 ..

- If target token isn't in the embedding, **removed it** and **looked at the next token**.
- Search the token until it exist in embeddings.
- Increased epoch to **100**.
- Compare **window size 2, 3 ...** and over.
- When searching for words(trigram, bigram, unigram) in embeddings
 - ① Raw word.
 - ② Lemmatizing.
 - ③ Stemming.

Experiment setting

- Pattern-based

- Pipeline

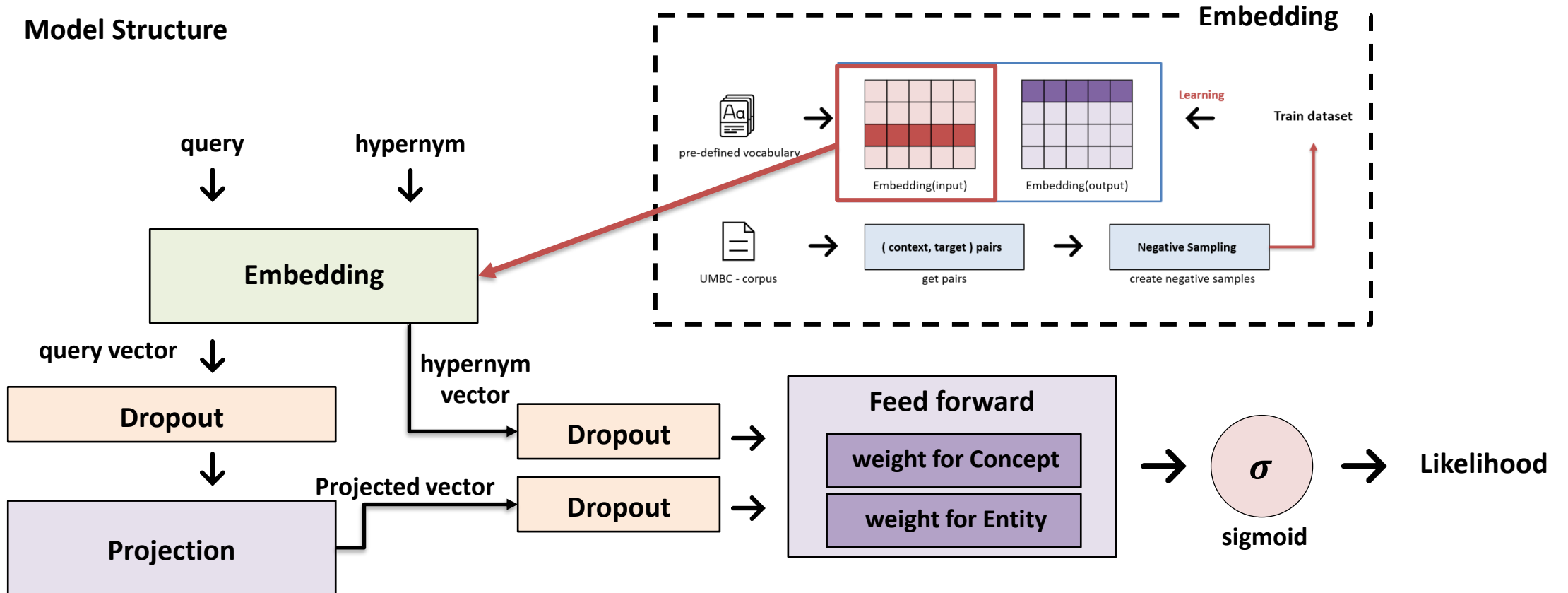


- From input query, find co-hyponyms of query and create co-hyponym set.
- Find hypernyms of co-hyponyms and calculate score based on frequency, embeddings.
- Extract 5 co-hyponyms first, get top 15 hypernyms by ranking hypernym of co-hyponym set.

Experiment setting

- Projection Learning

- Model Structure



- **Dropout** is applied to the embedding **query** and **hypernym** and **projected vector**.
- Train **separate** logistic regression classifier(for **Concept**, for **Entity**).

Experiment setting

- Projection Learning

- Projection Matrix

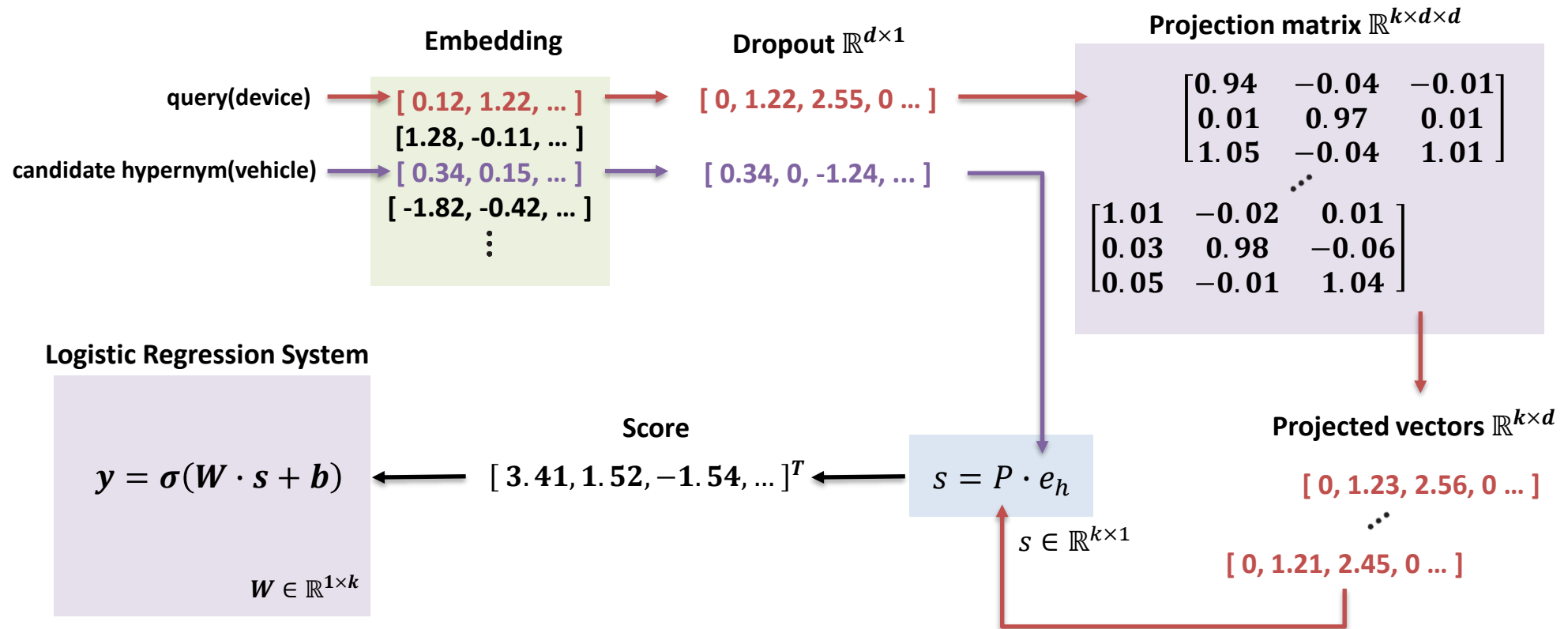
$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \xrightarrow{\text{noise}(0, 1/\text{dim})} \begin{pmatrix} 0.94 & -0.04 & -0.01 \\ 0.01 & 0.97 & 0.01 \\ 1.05 & -0.04 & 1.01 \\ \vdots & \vdots & \vdots \\ 1.01 & -0.02 & 0.01 \\ 0.03 & 0.98 & -0.06 \\ 0.05 & -0.01 & 1.04 \end{pmatrix}$$

(k, d, d) Identity Matrix
(k, d, d) Noised Matrix

- Initialize projection matrix to **identity matrix**, $\Phi \in \mathbb{R}^{k \times d \times d}$
- It means ***k* random copies** of the embedding vector.
- Apply **random noise** from the normal distribution $N(0, \frac{1}{\text{dim}})$

Experiment setting

- Projection Learning



- **Dropout** is applied to the embedding **query** and **hypernym** and **projected vector**.
- Train **separate** logistic regression classifier(for **Concept**, for **Entity**).

Experiment setting

- Projection Learning

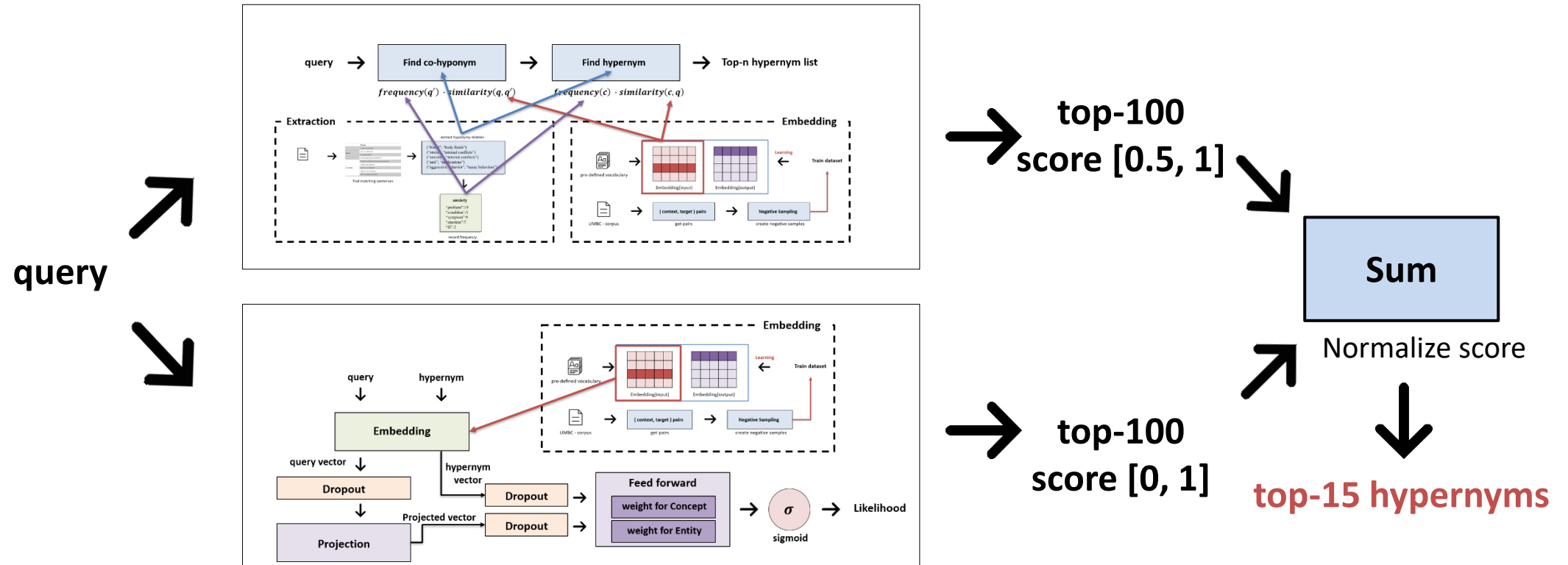
- Model settings

hyperparameter	setting
embedding dimension	100
negative samples	10
batch size	32
learning rate	0.0001
epoch	1000
patience (for MAP)	200

- Embeddings are trained by 408 corpus.
- Use **gradient clipping** during training.
- Model's optimizer is **Adam** and Projection's optimizer is **SGD**.
- Check validation MAP each iteration

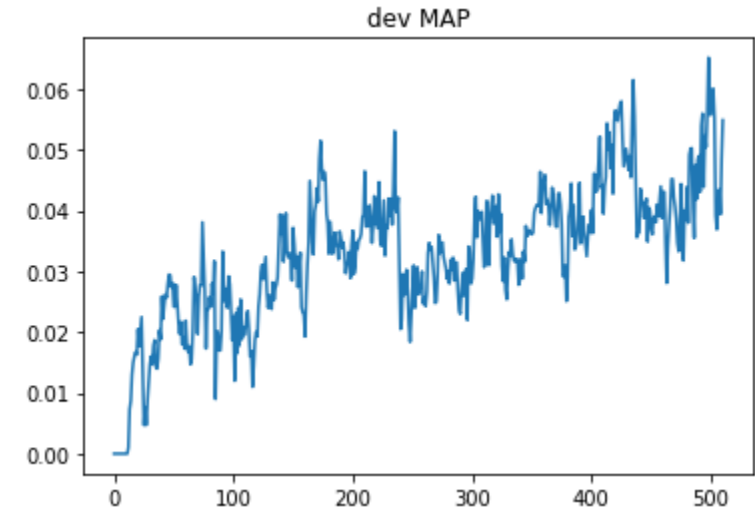
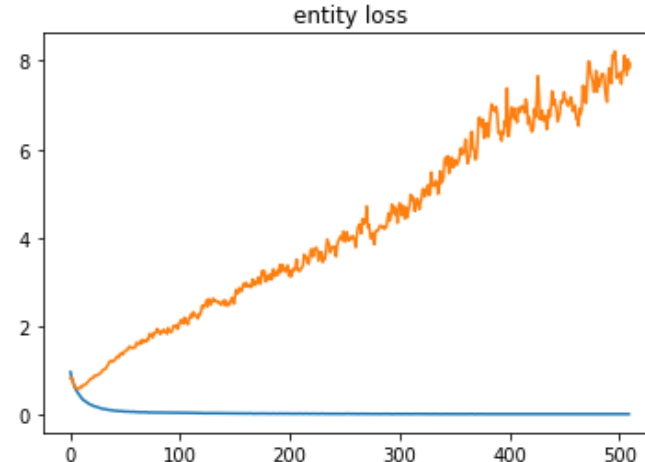
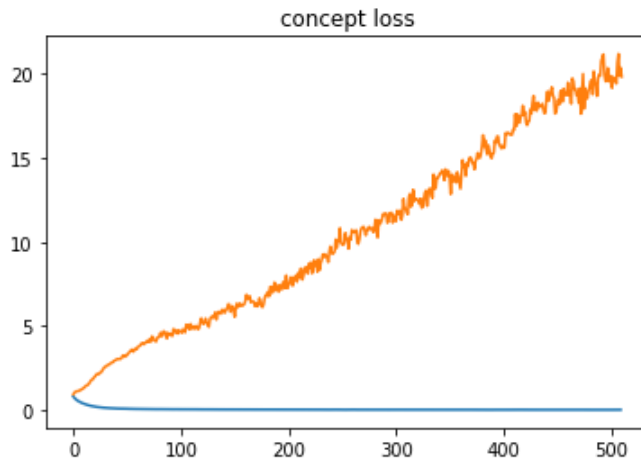
Experiment setting

- **Hybrid Model**
 - **Model Pipeline**



Experiment result

- Loss for Projection Learning



- Concept, Entity Projection models are trained by using **negative sampling**.
- Validation dataset is only **positive sample**.
- Check MAP of the validation dataset and stop training if the MAP does not increase for patience.

Experiment result

- Pattern-based

Model	MAP	MRR	P@1
CRIM	7.36	15.44	10.73
Before Unsupervised	5.94	12.63	9.00
Unsupervised – win2 - 10epoch	3.85	8.83	6.13
Unsupervised – win2 - 20epoch	4.09	9.24	6.73
Unsupervised – win2 - 30epoch	4.23	9.29	6.93
Unsupervised – win2 - 40epoch	4.25	9.43	7.06
Unsupervised – win2 - 50epoch	4.34	9.67	7.20
Unsupervised – win2 - 60epoch	4.50	9.92	7.33
Unsupervised – win2 - 70epoch	4.46	9.89	7.33
Unsupervised – win2 - 80epoch	4.47	9.89	7.33
Unsupervised – win2 - 90epoch			
Unsupervised – win2 - 100epoch			

Experiment result

- Projection Learning

Model	MAP	MRR	P@1
CRIM	19.11	34.99	28.80
Supervised-1	1.98	5.01	4.06
Supervised-100	3.06	7.81	6.13
Supervised-200	2.69	6.49	4.60
Supervised-300	2.99	7.17	4.93
Supervised-all	3.08	7.69	5.93
Supervised all corpora per 1 epoch	2.31	5.01	3.33
Supervised-win2-50epoch	7.95	19.04	16.06

Experiment result

- Hybrid

Model	MAP	MRR	P@1
CRIM	19.54	35.94	29.67
Hybrid-before	4.94	11.62	9.26
Hybrid-win2-50epoch	8.12	18.26	15.06
Supervised-win2-50epoch	7.95	19.04	16.06
Unsupervised – win2 - 60epoch	4.50	9.92	7.33

Issue & Future

- Issue

- Low MAP, MRR

- Increase window size for training embeddings.
 - Use pre-trained embeddings for training.

Issue & Future

- last week

- Checking overfitting.
- RMM(Recurrent Mapping Model)
- Embedding re-train

- Future works

- RMM(Recurrent Mapping Model)
- Use pre-trained embeddings.
- More experiments.

Thank you