

Hypernym discovery : CRIM experiments(4)

CRIM 구현 및 실험

Geonil Yun,

rjsdlf45@gmail.com

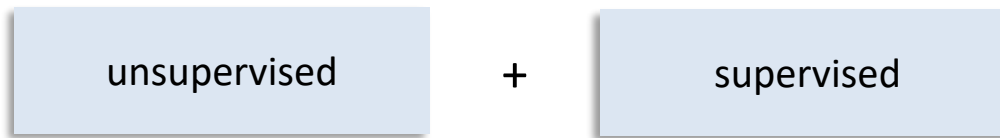
2021. 10. 19.



Introduction

- **CRIM – Hybrid Approach**

- Combination of two approaches



- Unsupervised : Pattern-based approach
- Supervised : Projection learning approach

- Dataset – SemEval2018 (general, medical, music)

- 2 query types, concept and named entity

Introduction

- **SemEval2018-Task9 : Hypernym discovery**

- 5 subtasks in SemEval2018-Task9

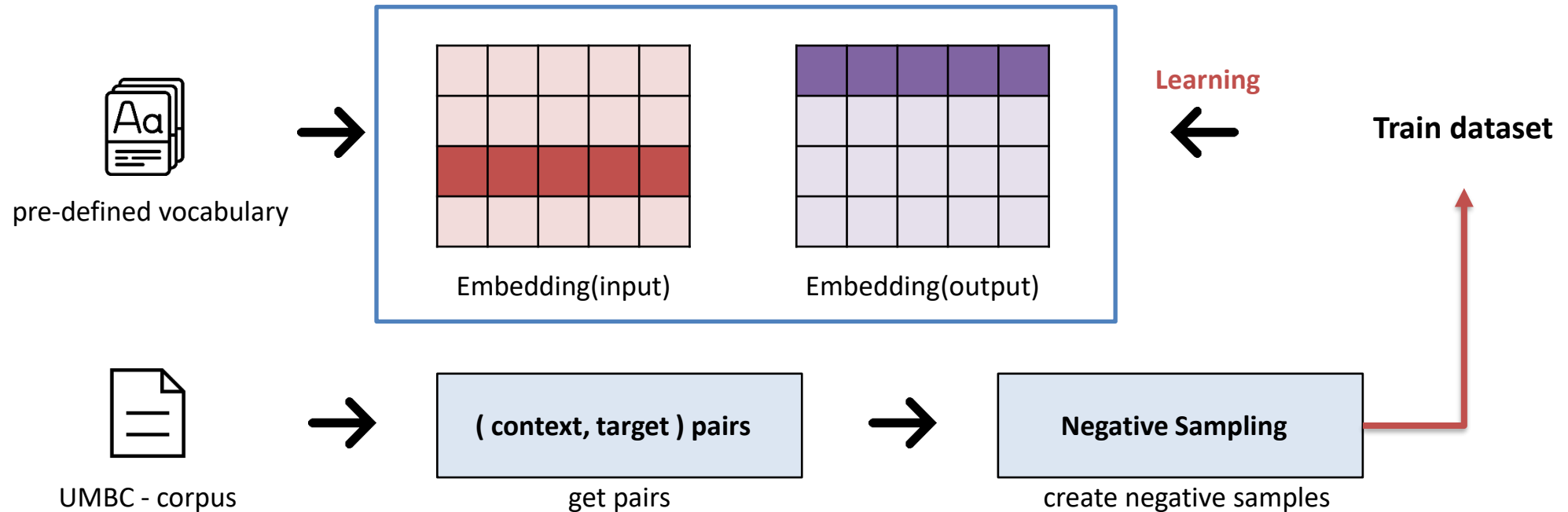
Task	General-purpose			Domain-specific	
	General			Medical	Music
Subtask	1A	1B	1C	2A	2B
Language	English	Italian	Spanish	English	English
Words	3B	3B	1.8B	130M	100M

- Find candidate hypernyms in pre-provided corpus.

Experiment setting

- Word Embeddings

- Skip-gram model structure



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	10
window size	2

before 1,024 , divergence



- Embeddings are trained by 408 corpus.
- Compare each embedding trained with 1, 100, 200, 300, 408 corpus.

Experiment setting

- **Pattern-based**

- Used patterns in CRIM

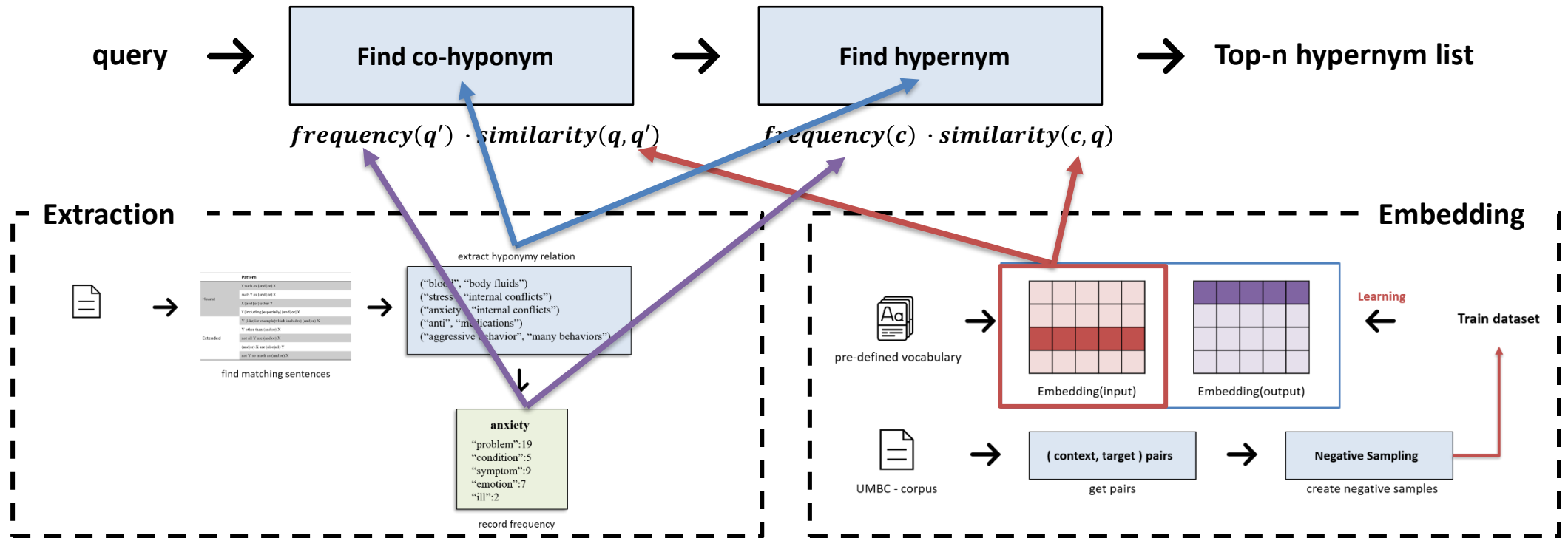
Pattern	
Co-hyponym	$X_1, X_2, \dots$ and X_n
Pattern	
Hearst	Y such as (and or) X
	such Y as (and or) X
	X (and or) other Y
	Y (including especially) (and or) X
	Y (like for example which includes) (and or) X
Extended	Y other than (and or) X
	not all Y are (and or) X
	(and or) X are (also all) Y
	not Y so much as (and or) X

- (X, Y) pair is (hyponym, hypernym) “is-a” relation.
- Use these patterns to find relation from corpus.

Experiment setting

- Pattern-based

- Pipeline

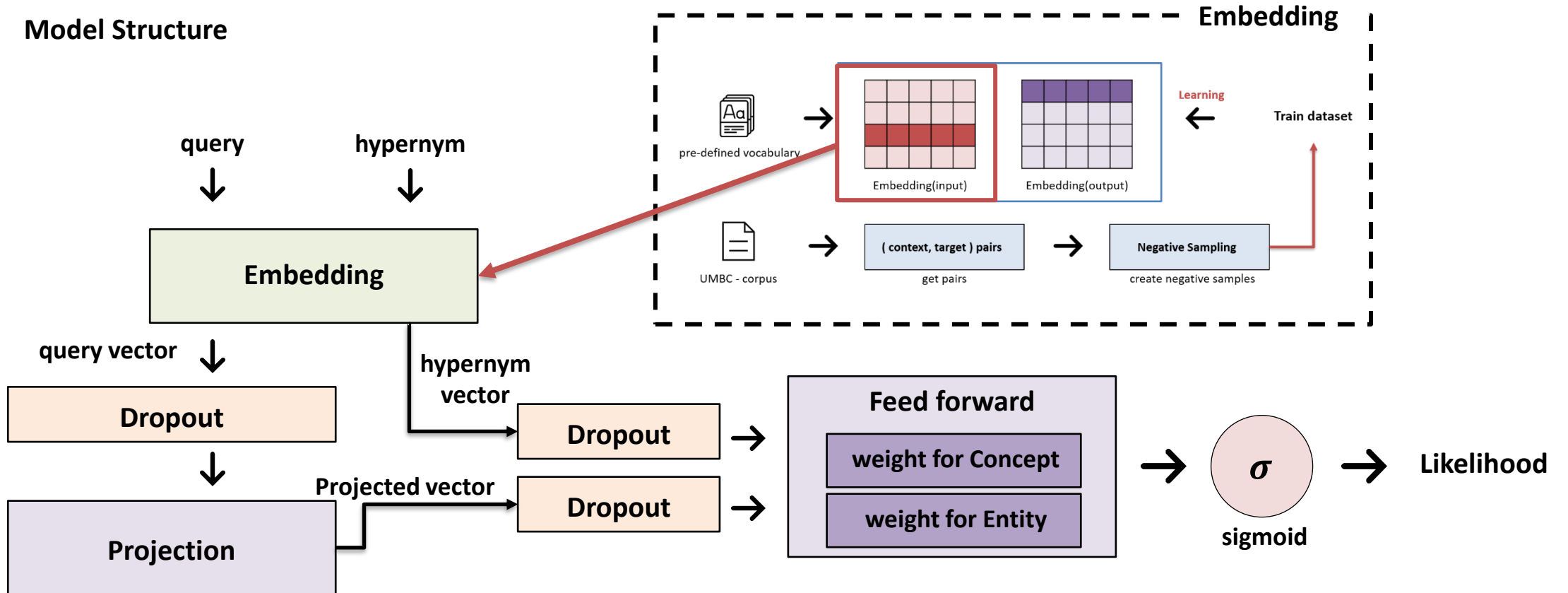


- From input query, find co-hyponyms of query and create co-hyponym set.
- Find hypernyms of co-hyponyms and calculate score based on frequency, embeddings.
- Extract 5 co-hyponyms first, get top 15 hypernyms by ranking hypernym of co-hyponym set.

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

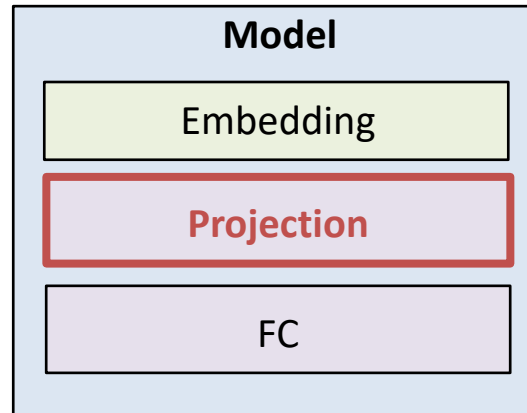
- Projection Learning

- Projection Matrix

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

(k, d, d) Identity Matrix

Initialize
➔



Learning
← Dataset (hyper-hypo) pairs

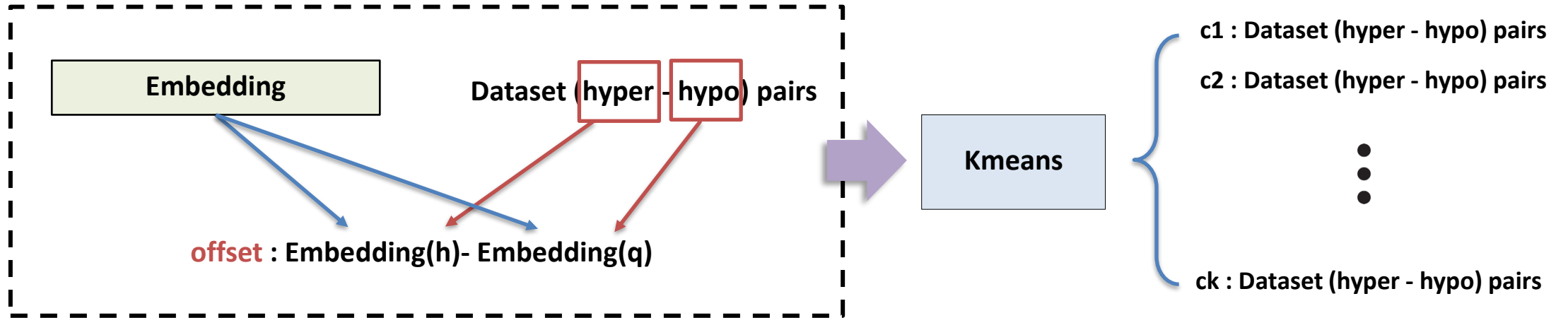
Output
➔ Bad results

- In projection layer, the projection matrix is initialized to a **(k x d x d) identity matrix**.
- After training the model, the output was not good.

Experiment setting

- Projection Learning

- Projection Matrix



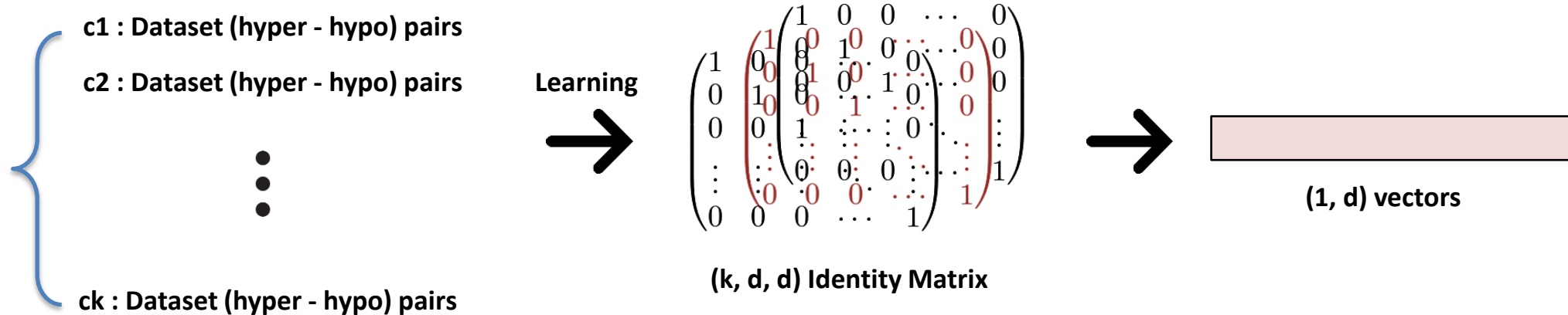
- First, the projection matrix was trained and then applied to the model.
- When training the projection matrix, ***k*** is **kernel size**.
- Produce ***k*** – **groups** by using **kmeans** Algorithm.

Experiment setting

- Projection Learning

- Projection Matrix

$$\Phi_k^* = \operatorname{argmin}_{\Phi_k} \frac{1}{N_k} \sum_{(x,y) \in C_k} \|\Phi_k x - y\|^2$$



- When training cluster k^{th} dataset, only use k^{th} dimension of projection matrix.
- Update Projection Matrix with MSE, SGD

Experiment setting

- Projection Learning

- Model settings

For Model	
hyperparameter	setting
embedding dimension	100
negative samples	10
batch size	32
learning rate	0.0001
epoch	100

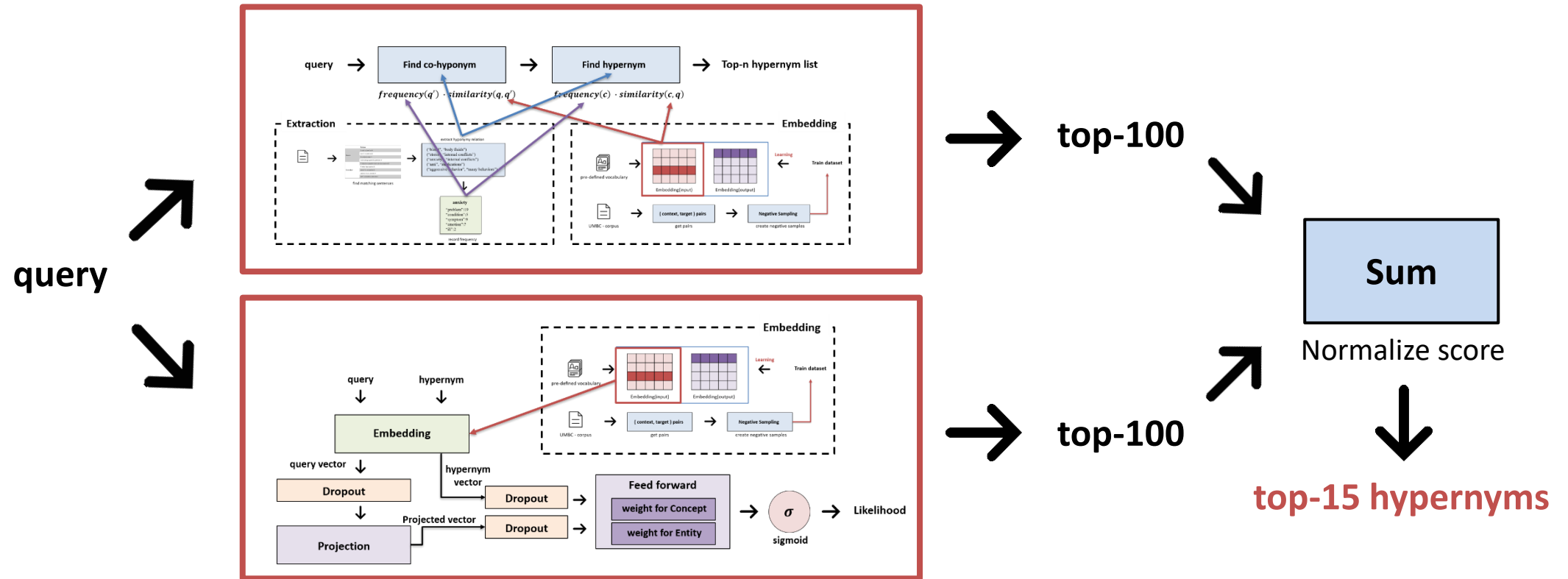
For Projection Matrix	
hyperparameter	setting
embedding dimension	100
batch size	32
learning rate	0.1
epoch	100

- Embeddings are trained by 408 corpus.
- Compare each embedding trained with 1, 100, 200, 300, 408 corpus.
- Use **gradient clipping** during training.
- Model's optimizer is **Adam** and Projection's optimizer is **SGD**.

Experiment setting

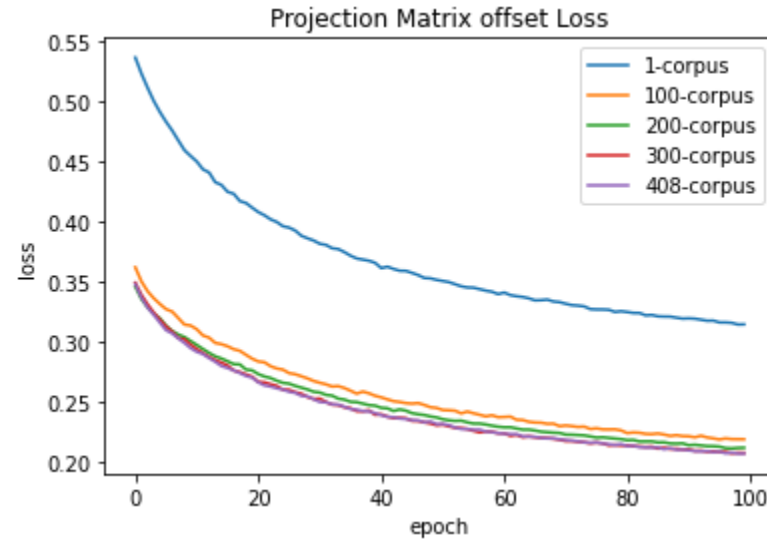
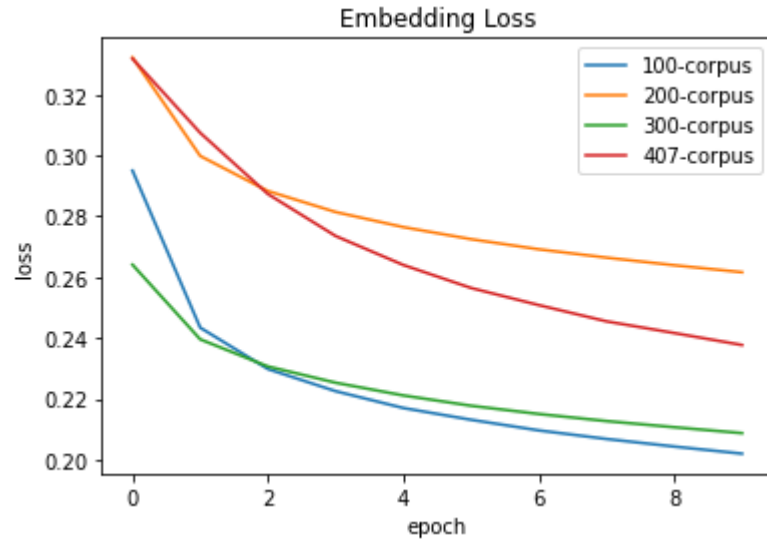
- Hybrid Model

- Model Pipeline



Experiment result

- Training graph



- Embedding loss decreased as more corpus were trained.
- Projection Matrix loss decreased.

Experiment result

- Pattern-based

Model	MAP	MRR	P@1
CRIM	7.36	15.44	10.73
Unsupervised-1	2.25	5.52	3.73
Unsupervised-100	4.52	10.01	7.33
Unsupervised-200	4.69	10.16	7.33
Unsupervised-300	4.81	10.15	7.13
Unsupervised-all	5.16	11.01	7.86

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

Experiment result

- Hybrid

Model	MAP	MRR	P@1
CRIM	7.36	15.44	10.73
Hybrid-1			
Hybrid-10			
Hybrid-20			
Hybrid-30			
Hybrid-40			

Issue & Future

- last week

- Training with whole dataset.
- Evaluation hybrid model.
- Check overfitting.

- Future works

- Evaluation hybrid model.
- Checking overfitting.
- Compare with other settings. (embedding epoch)
- RMM(Recurrent Mapping Model)

Thank you