

Hypernym discovery : CRIM experiments(5)

CRIM 구현 및 실험

Geonil Yun,

rjsdlf45@gmail.com

2021. 10. 26.



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

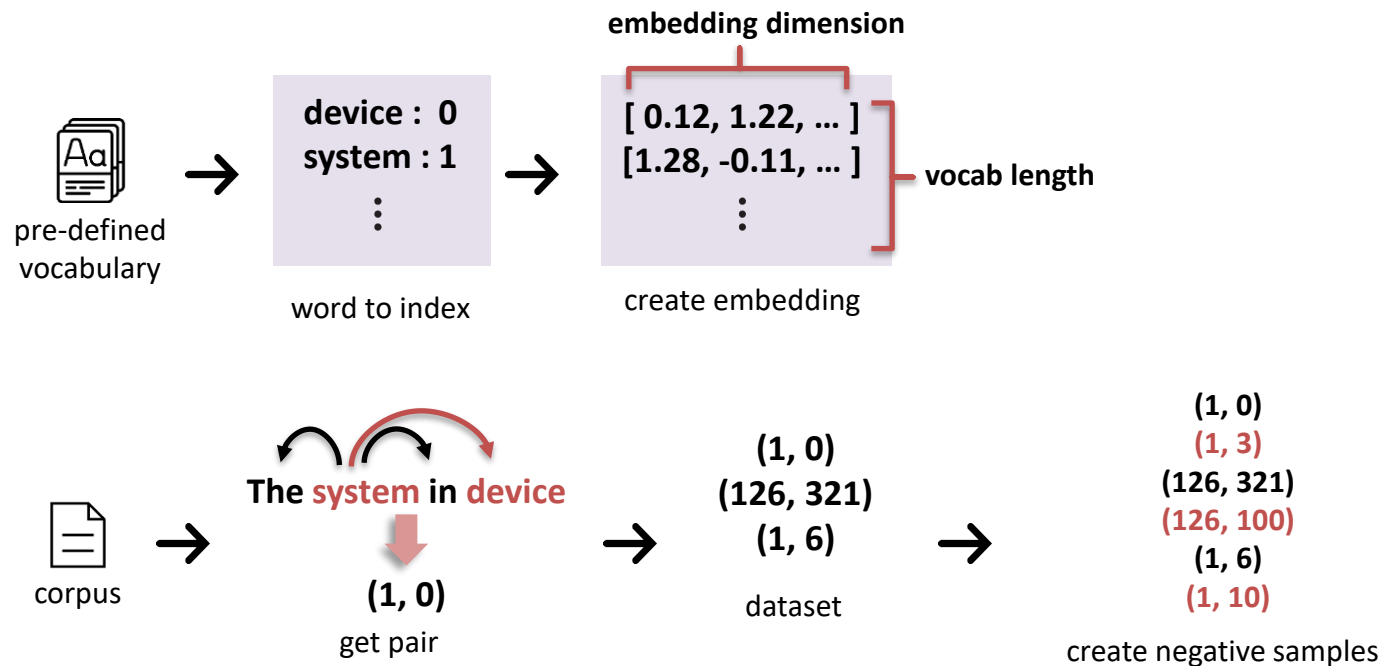
Task	General-purpose			Domain-specific	
	General	Medical	Music		
Subtask	1A	1B	1C	2A	2B
Language	English	Italian	Spanish	English	English
Words	3B	3B	1.8B	130M	100M

- Find candidate hypernyms in pre-provided corpus.

Experiment setting

- Word Embeddings

- Skip-gram model structure



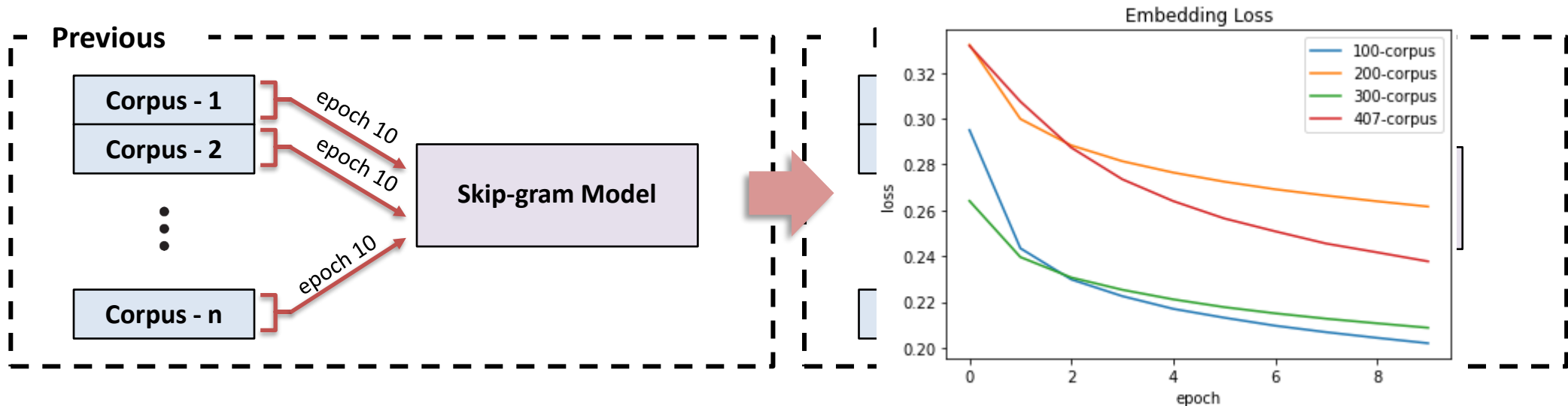
hyperparameter	setting
embedding dimension	100
negative samples	15
batch size	300,000
learning rate	0.0001
epoch	10
window size	2

- Create word to index dictionary for representation the word to integer.
- Create two embeddings for the input words and the output words.
- Build dataset from corpus by using the word to index dictionary, and use negative sampling.

Experiment setting

- Word Embeddings

- Skip-gram model training



- Previously, the model trained **10 epochs per corpus**.
- Each corpus has **different topics**.
- When training the last corpus, the **effect of the first corpus on embeddings** is reduced.
- So, Train **all corpora per 1 epoch**.

Experiment setting

- **Pattern-based**

- Used patterns in CRIM

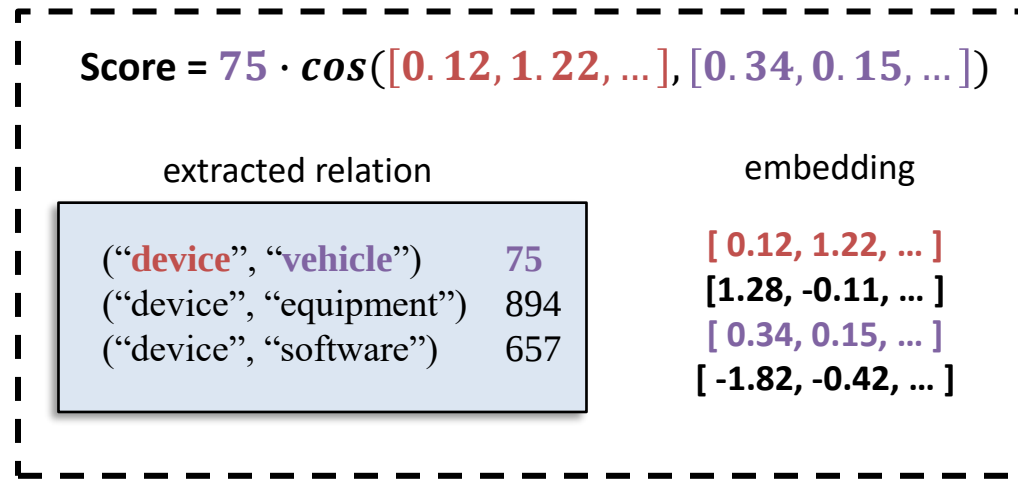
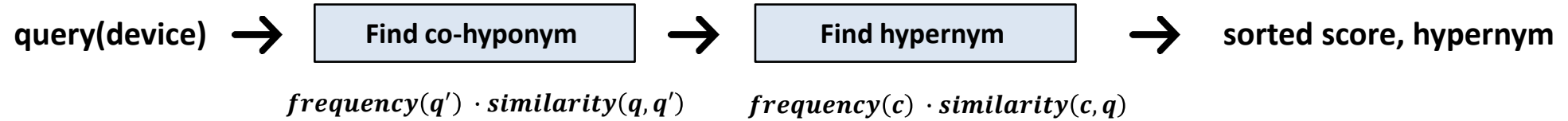
Pattern	
Co-hyponym	$X_1, X_2, \dots$ and X_n
Pattern	
Hearst	Y such as (and or) X
	such Y as (and or) X
	X (and or) other Y
	Y (including especially) (and or) X
	Y (like for example which includes) (and or) X
Extended	Y other than (and or) X
	not all Y are (and or) X
	(and or) X are (also all) Y
	not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

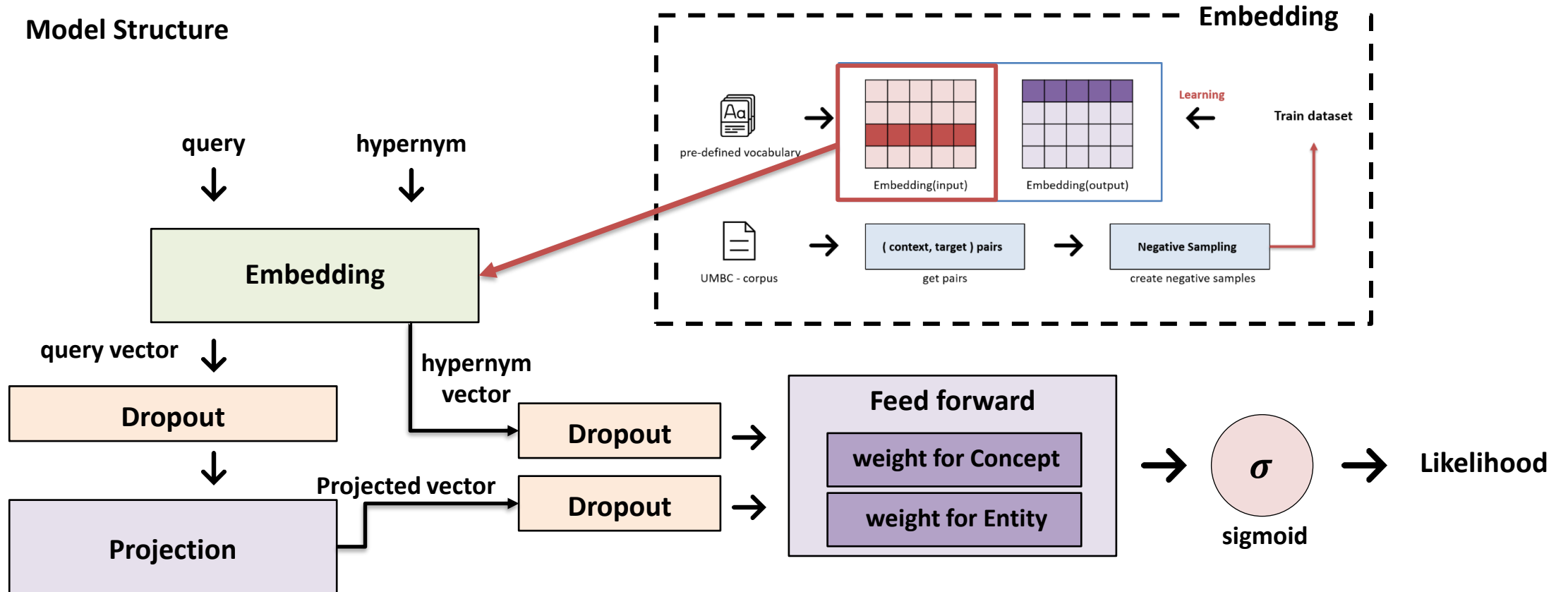


- From input query, find co-hyponyms of query and create co-hyponym set.
- Find hypernyms of co-hyponyms and calculate score based on frequency, embeddings.
- Extract 5 co-hyponyms first, get top 15 hypernyms by ranking hypernym of co-hyponym set.

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

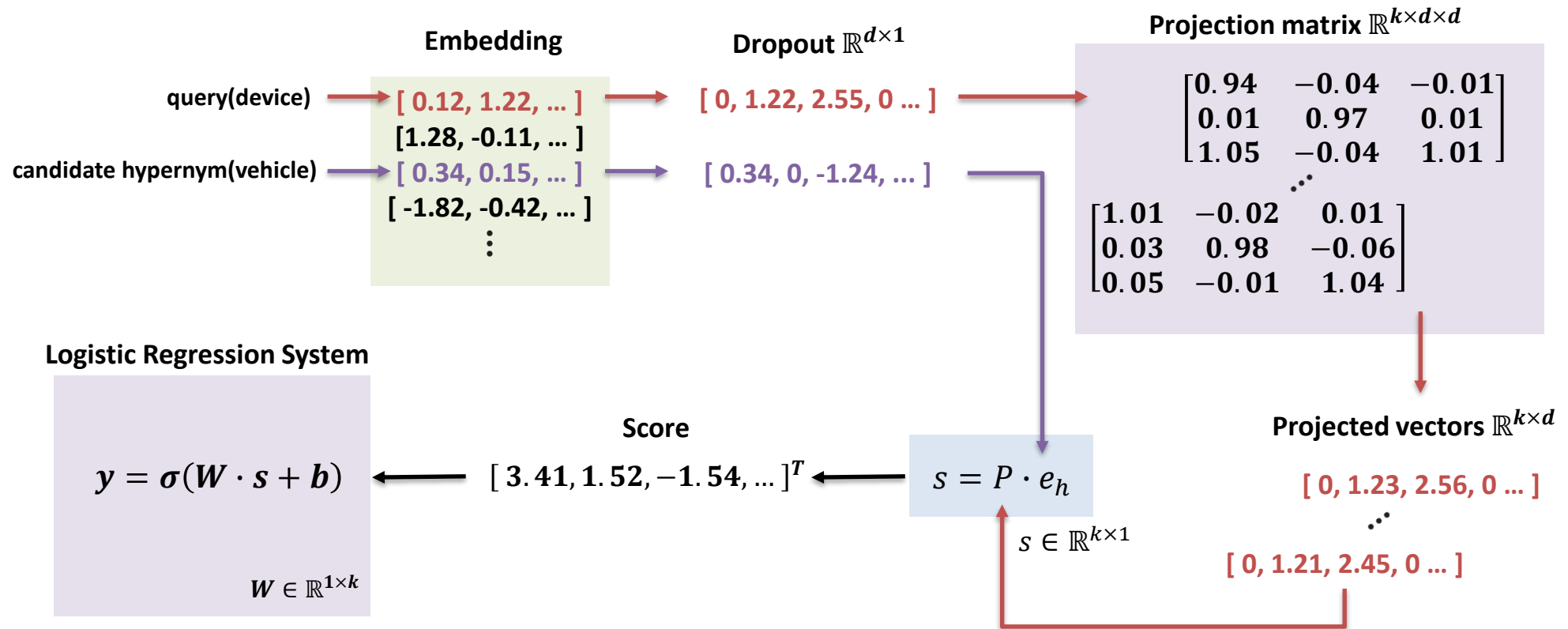
$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \xrightarrow{\text{noise}(0, 1/\text{dim})} \begin{pmatrix} 0.94 & -0.04 & -0.01 \\ 0.01 & 0.97 & 0.01 \\ 1.05 & -0.04 & 1.01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

(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

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

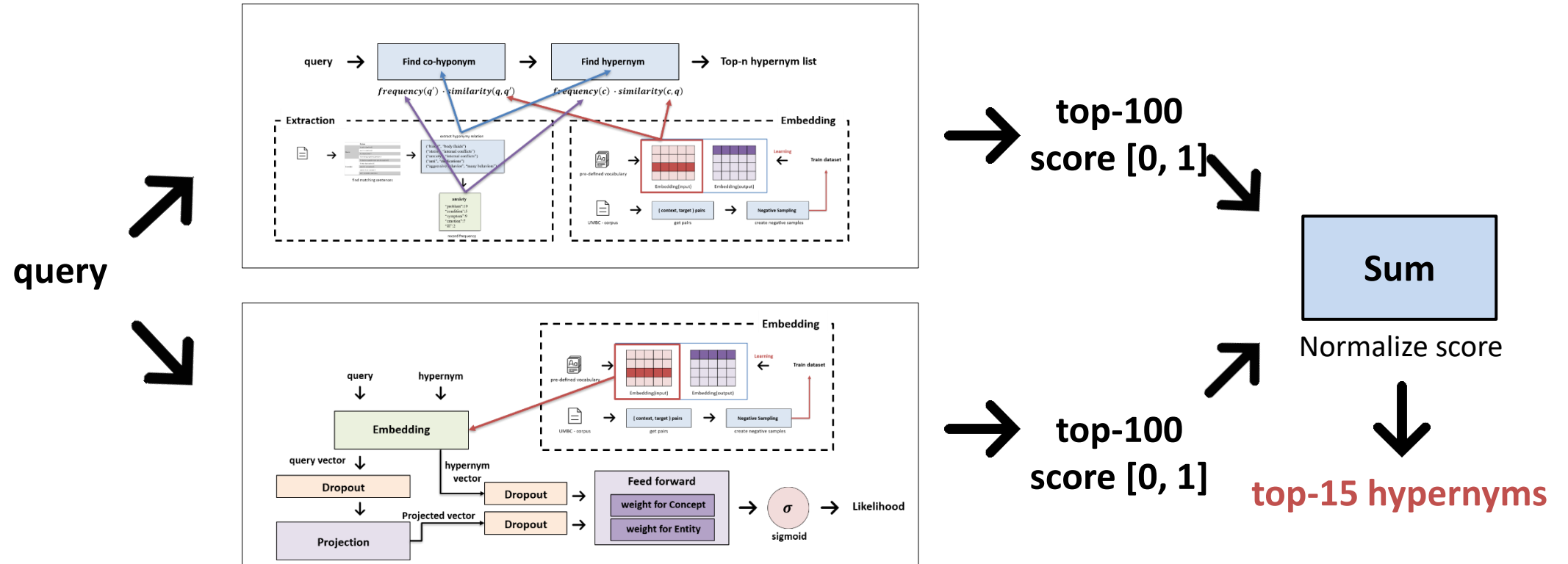
For Projection Matrix	
hyperparameter	setting
embedding dimension	100
batch size	32
learning rate	0.1
epoch	100

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

Experiment setting

- Hybrid Model

- Model Pipeline



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
Unsupervised all corpora per 1 epoch	5.94	12.63	9.00

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

Experiment result

- Hybrid

Model	MAP	MRR	P@1
CRIM	19.54	35.94	29.67
Hybrid	4.94	11.62	9.26
unsupervised	5.94	12.63	9.00
supervised	2.31	5.01	3.33

Issue & Future

- Issue

- Several queries are not in word2idx dictionary
- Empty results are reduce performance.
- In paper, they also train queries embedding.

contre	streich	dead body	unhumorous	memory trace	ma	
title respect	marking	current air	constituent	electronic co		
imaginary place	motion picture		sandstonelithostratigraphic unit		ves	
mechanical assembly	system program		heavier-than-air craft		mo	
single-valued function		app	service	channel	transmission	tec
sportsperson	person	chief	picture	political leader	tropical cycl	
literary work	control	film	specification	leader	illness	pla
fluid	social event		structural member	monarch	foodstuff	liquid state
person	physical phenomenon		peninsula state	currency sign	mo	
technical specification		natural virtue	wrongful act	actus reus		

Issue & Future

- last week

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

- Future works

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

Thank you