

Hypernym discovery : CRIM experiments

CRIM 구현 및 실험

Geonil Yun,

rjsdlf45@gmail.com

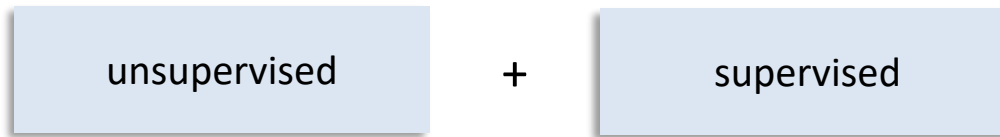
2021. 09. 28.



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

- **Pattern-based**

- Used patterns in CRIM

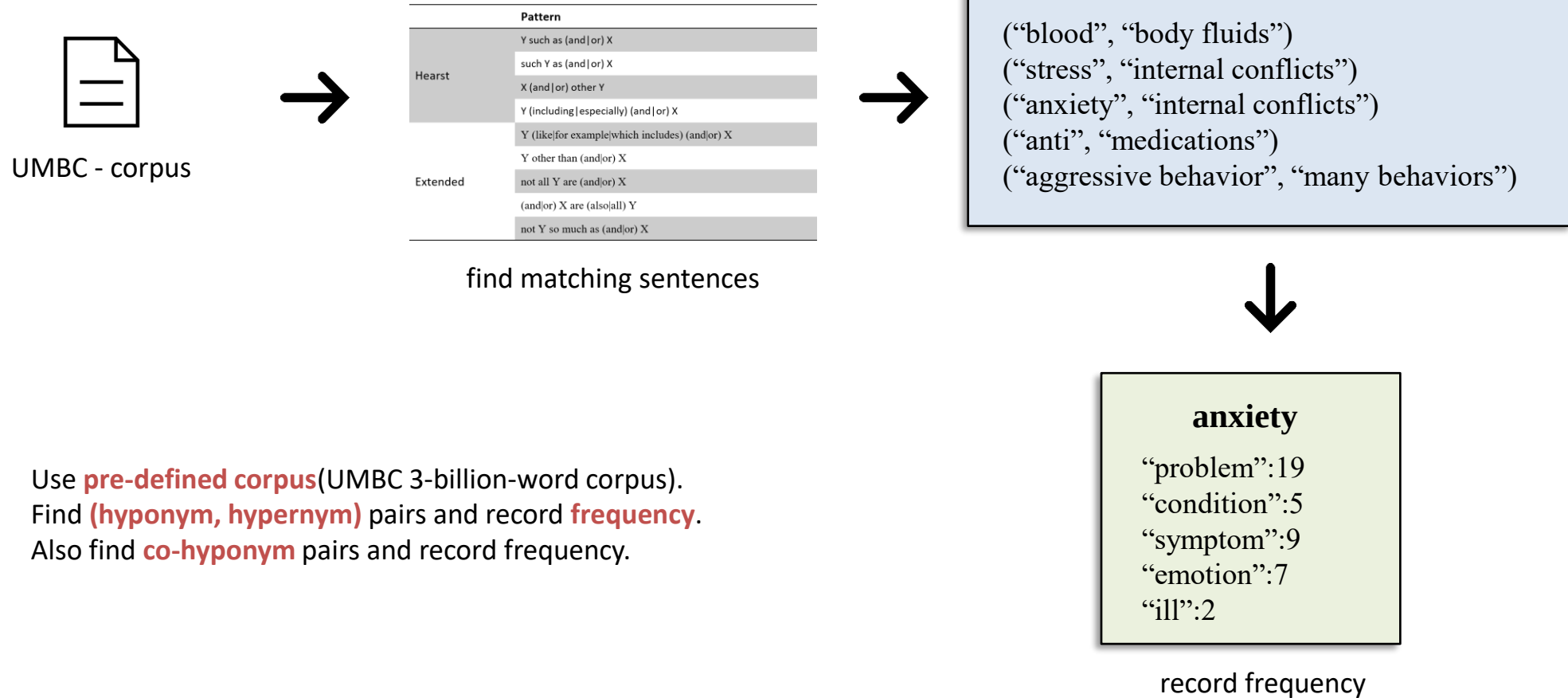
Pattern	
Co-hyponym	$X_1, X_2, \dots$ and X_n
Pattern	
Hearst	Y such as (and or) X
	such Y as (and or) X
	X (and or) other Y
	Y (including especially) (and or) X
	Y (like for example which includes) (and or) X
Extended	Y other than (and or) X
	not all Y are (and or) X
	(and or) X are (also all) Y
	not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

- Extraction hyponymy

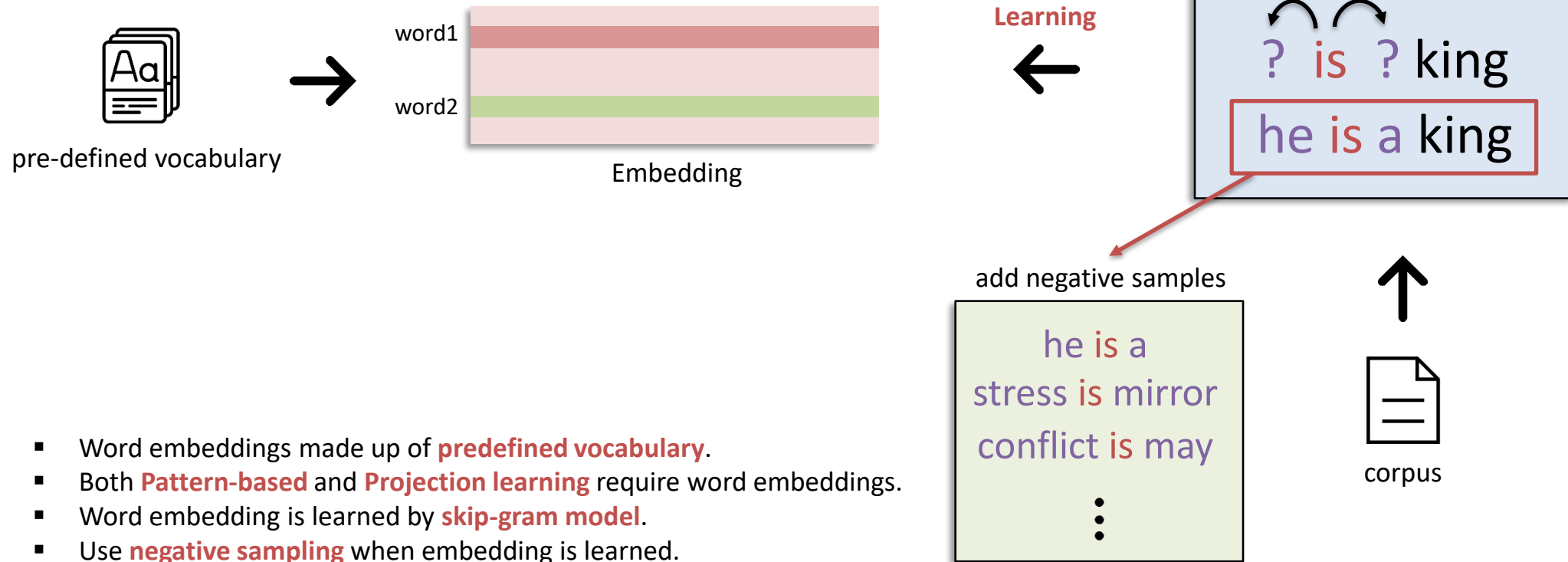


- Use **pre-defined corpus**(UMBC 3-billion-word corpus).
- Find (**hyponym, hypernym**) pairs and record **frequency**.
- Also find **co-hyponym** pairs and record frequency.

Experiment setting

- Word embeddings

- Skip-gram

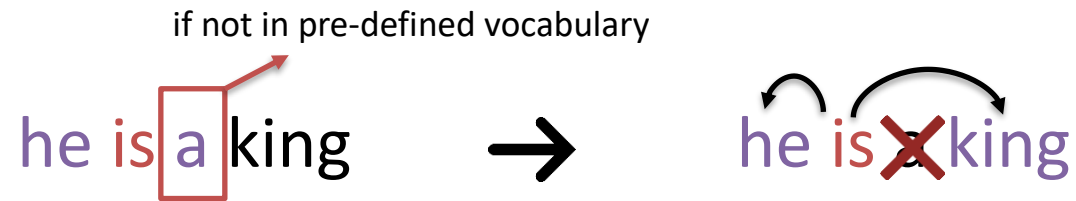


Issue & Future

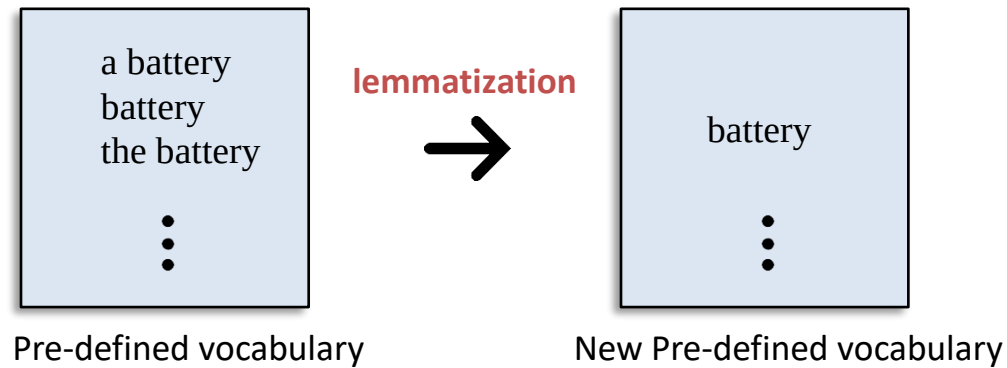
- ISSUE

- Learning embeddings only pre-defined vocabulary

- ① Delete OOV(Out-of-Vocabulary)



- ② Lemmatization pre-defined vocabulary



Issue & Future

- **Future works**
 - Extract co-hyponym, hyponymy from all files.
 - Learning embeddings from corpus.
 - Evaluate pattern-based approach by using embeddings and extracted patterns.
 - Implement projection learning model. (projection matrix)

Thank you