Neural Natural Logic Inference for Interpretable Question Answering

EMNLP 2021

Jihao Shi, Xiao Ding, Li Du, Ting Liu, and Bing Qin

2022.02.17

AILAB 조환희

Motivation

Open Domain QA의
Explainability 부족

↓

Natural Logic 으로 설명 가능하게 하겠다!

Neural Natural Logic Inference
for Interpretable Question Answering

Neural **Natural Logic Inference**
for Interpretable Question Answering

Background

Natural Logic

삼단논법

두 판단에서 그것들과는 다른 하나의 새로운 판단으로 이끄는 **추론 방법**

전제1	모든 사람은 죽는다.
전제2	소크라테스는 사람이다.
결론	소크라테스는 죽는다.

Natural Logic

형식화된 추론

← 문법적인 구조와 의미론적인 특징을 사용

드 모르간의 법칙

$$\backsim (p \vee q) \equiv \backsim p \wedge \backsim q$$

p	오리는 악어가 아니다.
q	고양이는 개가 아니다.
$(p \vee q)$	오리는 악어가 아니거나 고양이는 개가 아니다.
$\backsim (p \vee q)$	오리는 악어가 아니거나 고양이는 개가 아닌 것은 거짓이다.
$\backsim p \wedge \backsim q$	오리는 악어이고 고양이는 개다.

Background

Natural Logic

형식화된 추론

dog $\leq$ animal 라는 관계가 성립된다고 가정했을 때, ($\leq \approx \sqsubseteq$)

$$\frac{\begin{array}{c}\text{not every animal barks.}\\ \text{not every dog barks.}\end{array}}{\text{not every dog barks.} \leq \text{not every animal barks.}}$$

이런 관계를 표현한 것을 Monotonicity Calculus라 한다.

Background

Natural Logic

형식화된 추론

Monotonicity Calculus

“dog ≤ animal” 이고 “not every dog barks. ≤ not every animal barks.”

Monotonic

“dog ≤ animal” 이고 “every dog barks ≥ every animal barks.”

Antitonic

Natural Logic

형식화된 추론

Monotonicity Calculus

“dog $\leq$ animal” 이고 “not every dog barks. $\leq$ not every animal barks.”

Monotonic

“dog $\leq$ animal” 이고 “every dog barks $\geq$ every animal barks.”

Antitonic

더 general 하게 표현하는 것을 monotonic이라 한다.

Background

Natural Logic

형식화된 추론

Polarity

"dog ≤ animal" 이고 "not every dog barks. ≤ not every animal barks."

not every ↑dog barks.

upward monotonic

"dog ≤ animal" 이고 "every dog barks ≥ every animal barks."

every ↓dog barks.

downward monotonic

Background

Natural Logic

형식화된 추론

Polarity

All lexical items are positive polarity

some, several, or a few
– preserve polarity

like no, not, and all (in its first argument)
– reverse polarity.

"**no** cats eat ↓mice"
"mice in **no** cats do**n't** eat ↑mice"

Background

Natural Language Inference(NLI)

자연어 추론

Datasets : MNLI, SNLI, …

determining whether a "hypothesis" is true (entailment), false (contradiction), or undetermined (neutral) given a "premise".

Premise	A senior is waiting at the window of a restaurant that serves sandwiches
Relationship	entailment
Hypothesis	A person waits to be served his food

Natural Language Inference(NLI)

An extended model of natural logic, ICCS, 2004, Bill MacCartney and Christopher D. Manning

6가지의 lexical간 relation symbol

Monotonicity calculus는 entailment만 표현 가능했다.

symbol[5]	name	example	set theoretic definition[6]
$x \equiv y$	equivalence	*couch* $\equiv$ *sofa*	$x = y$
$x \sqsubset y$	forward entailment	*crow* $\sqsubset$ *bird*	$x \subset y$
$x \sqsupset y$	reverse entailment	*European* $\sqsupset$ *French*	$x \supset y$
$x \wedge y$	negation	*human* $\wedge$ *nonhuman*	$x \cap y = \emptyset \wedge x \cup y = U$
$x \mid y$	alternation	*cat* $\mid$ *dog*	$x \cap y = \emptyset \wedge x \cup y \neq U$
$x \smile y$	cover	*animal* $\smile$ *nonhuman*	$x \cap y \neq \emptyset \wedge x \cup y = U$
$x \# y$	independence	*hungry* $\#$ *hippo*	(all other cases)

Background

Natural Language Inference(NLI)

Natural Language Inference(NLI) task를

insertion, deletion, substitution 등으로 풀이

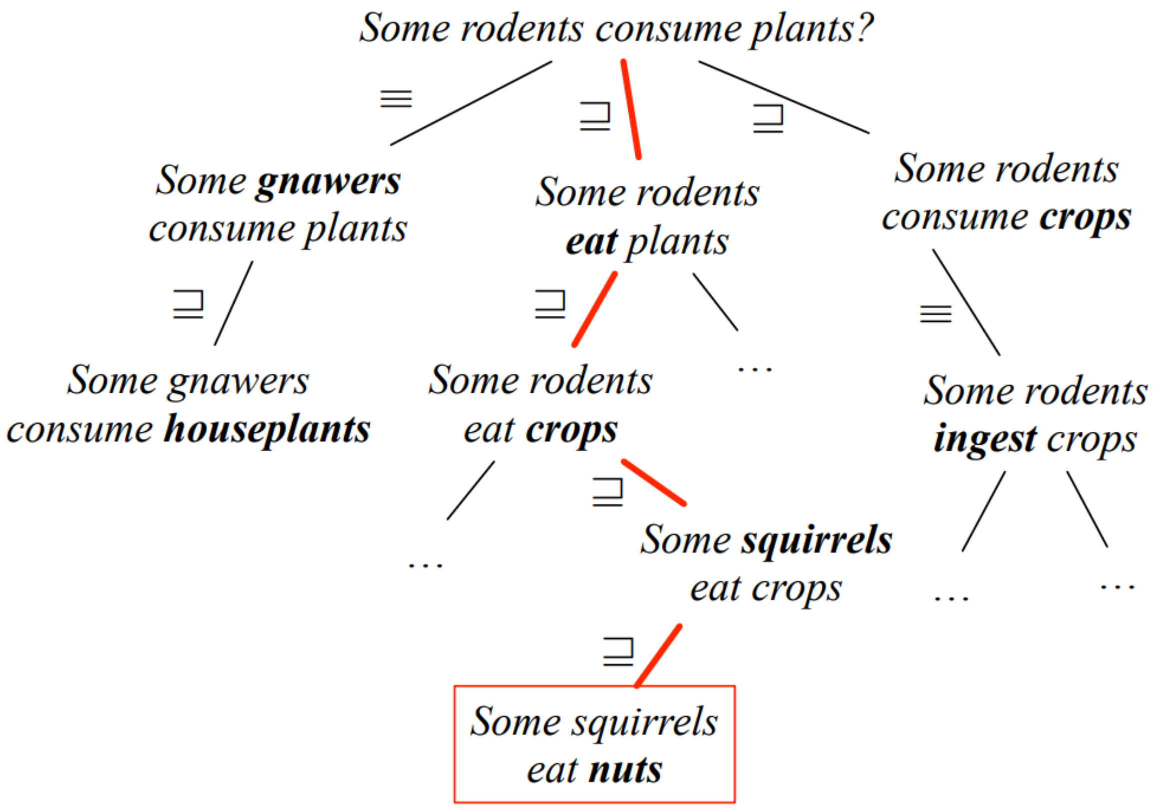
- 3개의 action은 앞선 6개의 relation으로 표현 가능

Substitution	crow⊏bird, sofa≡couch,
Deletion	red car ⊏ car, former student ∣ student
Insertion	Sing ⊐ sing off-key

Natural Language Inference(NLI)

Natural Language Inference(NLI) task를 insertion, deletion, substitution 등으로 풀이
(Upward monotone(general 한 관계가 유지되게) 방향으로 확장)

Premise	Some squirrels eat nuts.
Hypothesis	Some rodents consume plants.



QA as Natural Language Inference(NLI)

Premise가 true라면 hypothesis또한 true인 것

=Premise로 hypothesis를 추론할 수 있다.

≡Premise entails hypothesis

$\equiv p \vDash h$

premise	소크라테스는 사람이고 모든 사람은 죽는다.
hypothesis	소크라테스는 죽는다.

Problem Setup

QA as Textual Entailment

determine premise that entails one of four hypothesis

Example–1:
Question: The main function of a fish's fins is to help the fish _____.
(A) reproduce (B) see (C) breathe (D) move
Knowledge Base: . . . A fish has a flipper or fin that helps them swim. The dorsal fin can help to keep the fish stable in the water. . . .

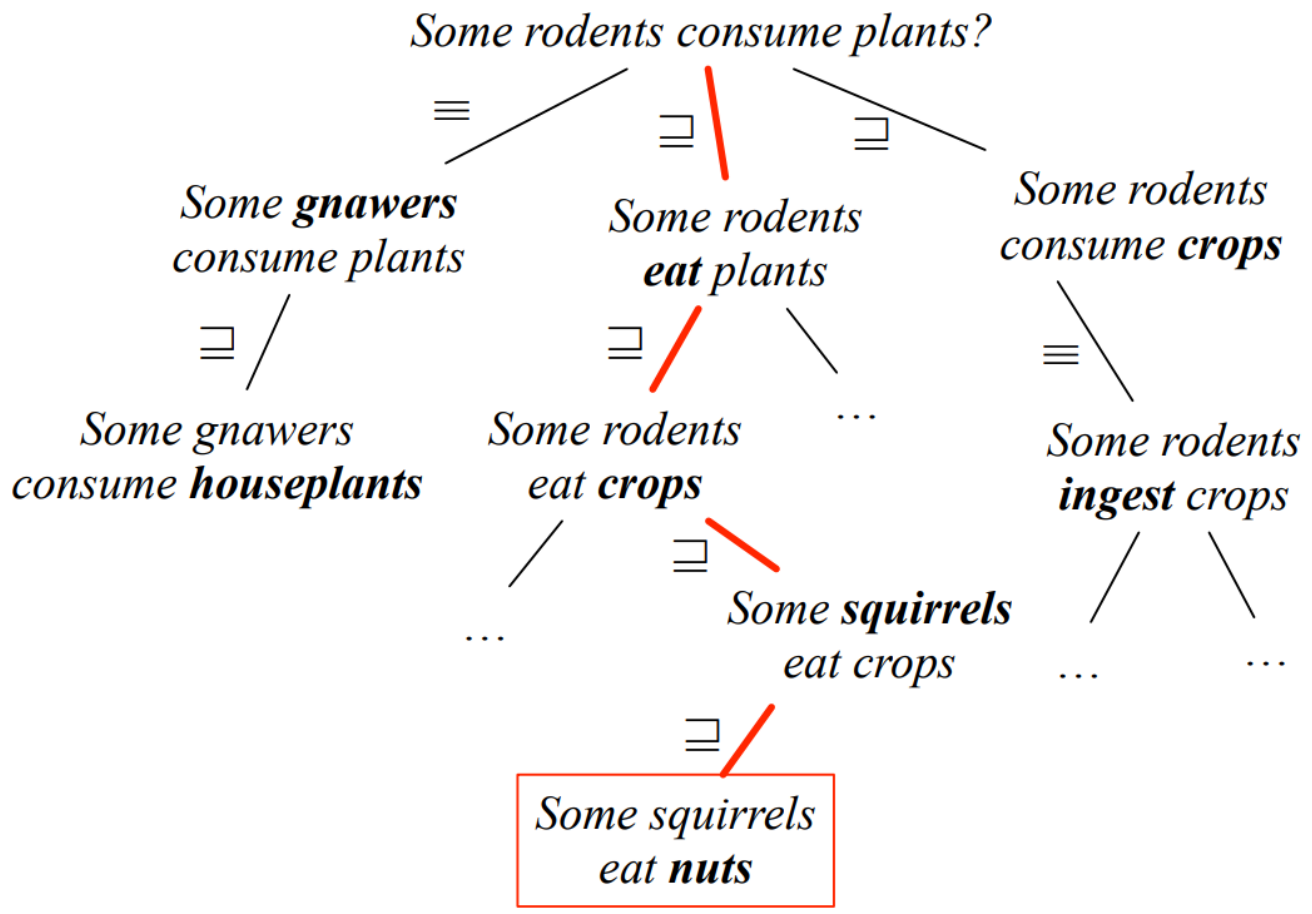
$H = \{h_1, h_2,, h_3,, h_4\}$: Q+A

h_1	The main function of a fish's fins is to help the fish reproduce.
h_2	The main function of a fish's fins is to help the fish see.
⋮	⋮

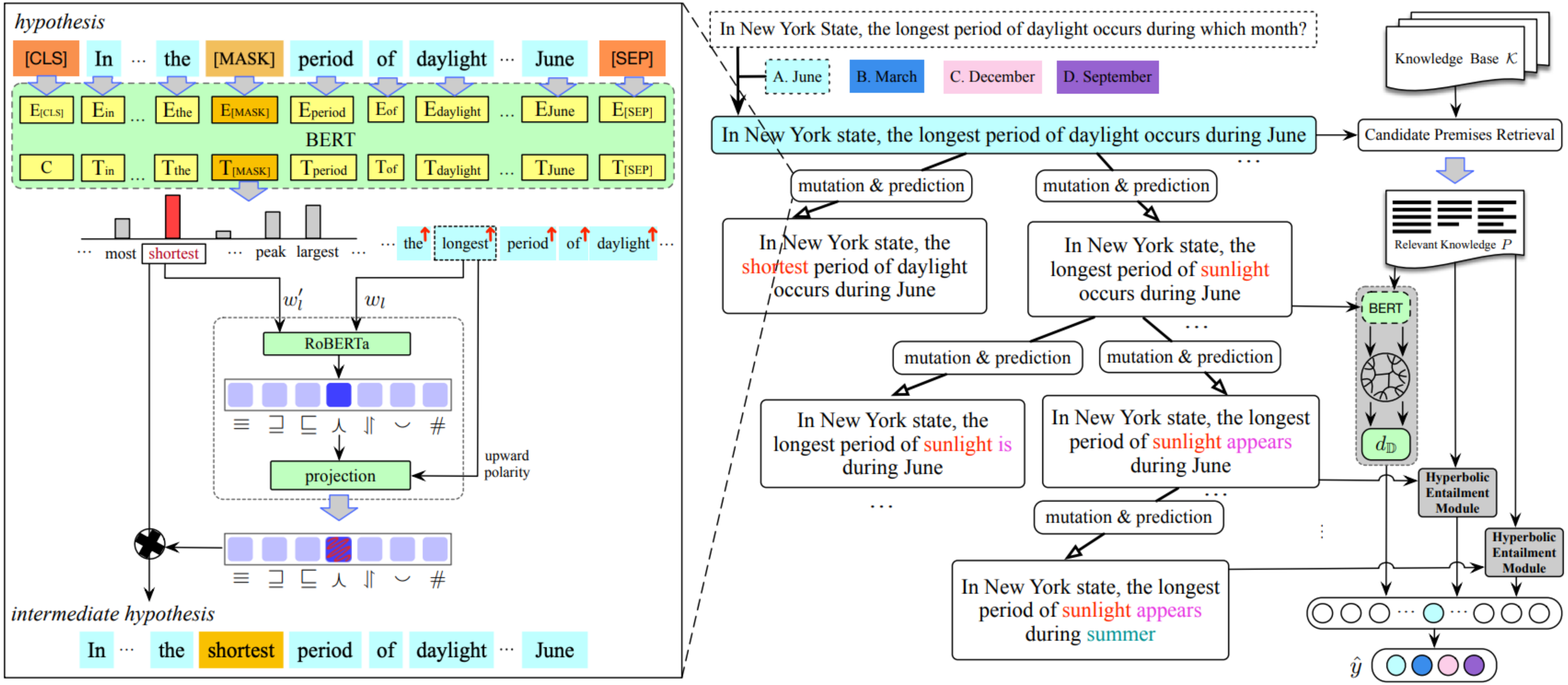
$P = \{p_1, p_2, \ldots\}$: relevant premise from KB

p_1	A fish has a flipper or fin that helps them swim.
p_2	The dorsal fin can help to keep the fish stable in the water
⋮	⋮

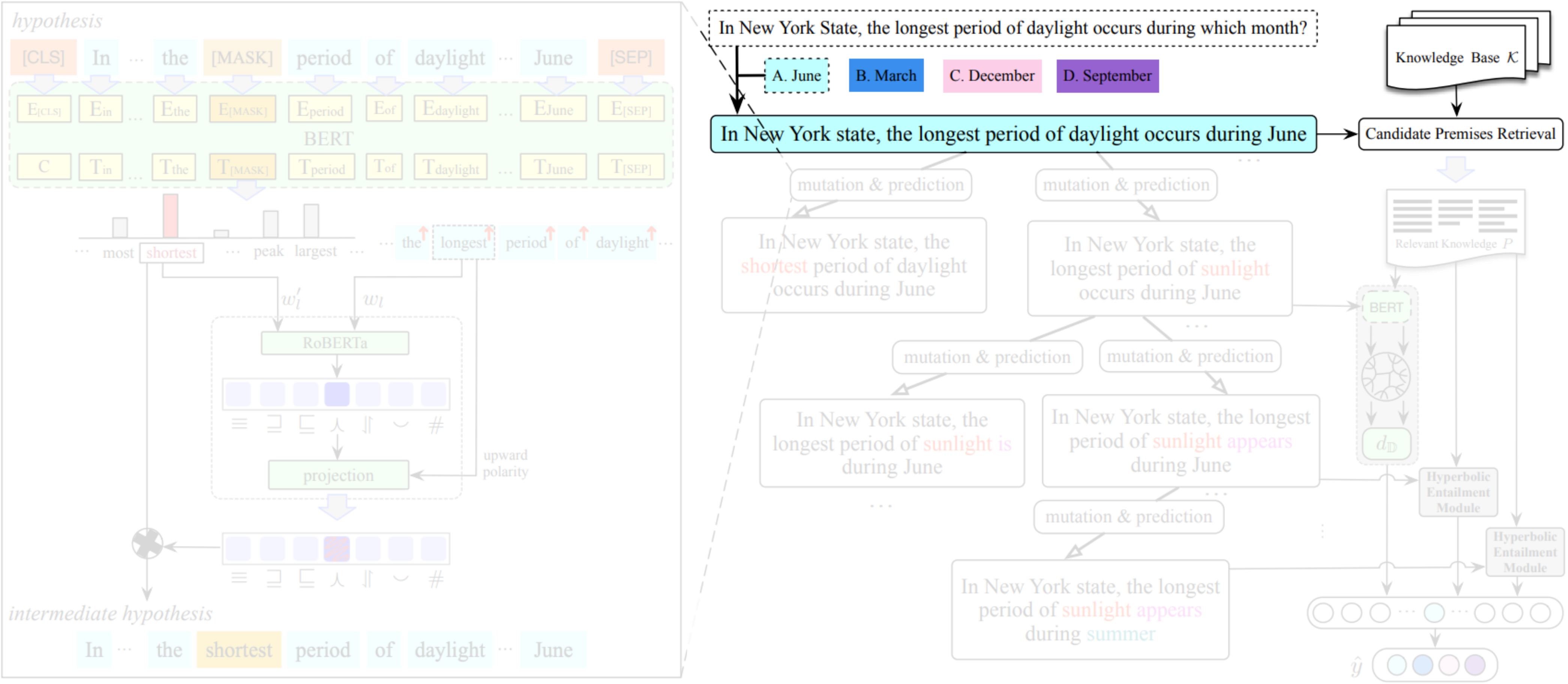
Hypothesis로부터 Premise를 Search



Architecture

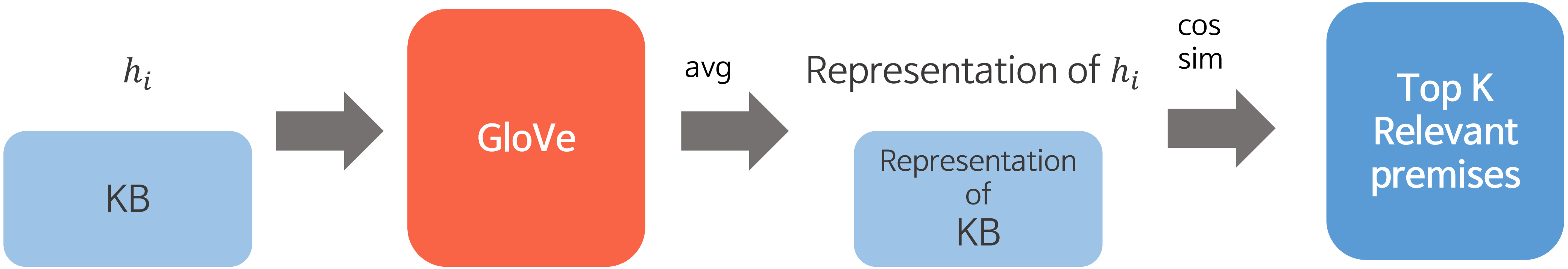


Candidate Premises Retrieval



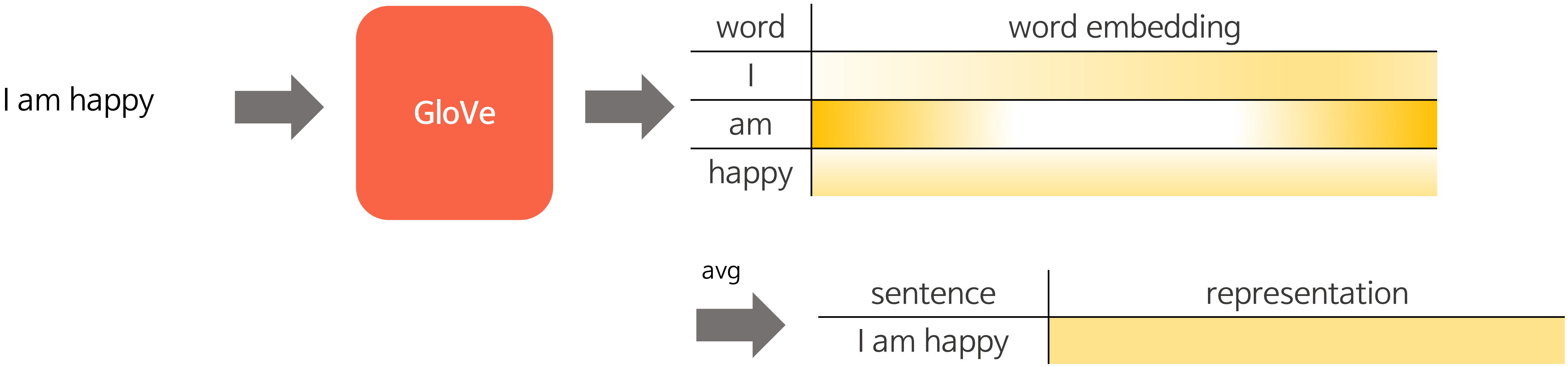
Candidate Premises Retrieval

1. representation of h_i and each p_j in KB by computing the average Glove word embeddings
2. calculate the cosine similarity between h_i and each p_j to find the top k relevant candidate premises



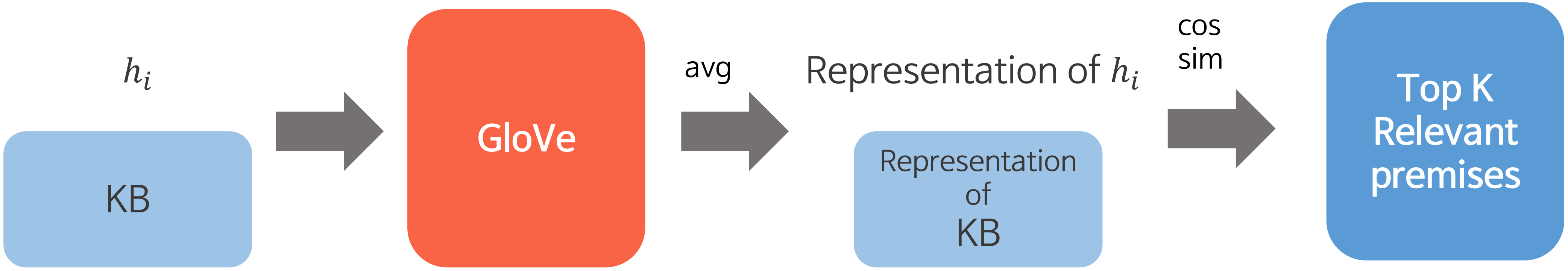
Candidate Premises Retrieval

1. representation of h_i and each p_j in KB by computing the average Glove word embeddings
2. calculate the cosine similarity between h_i and each p_j to find the top k relevant candidate premises

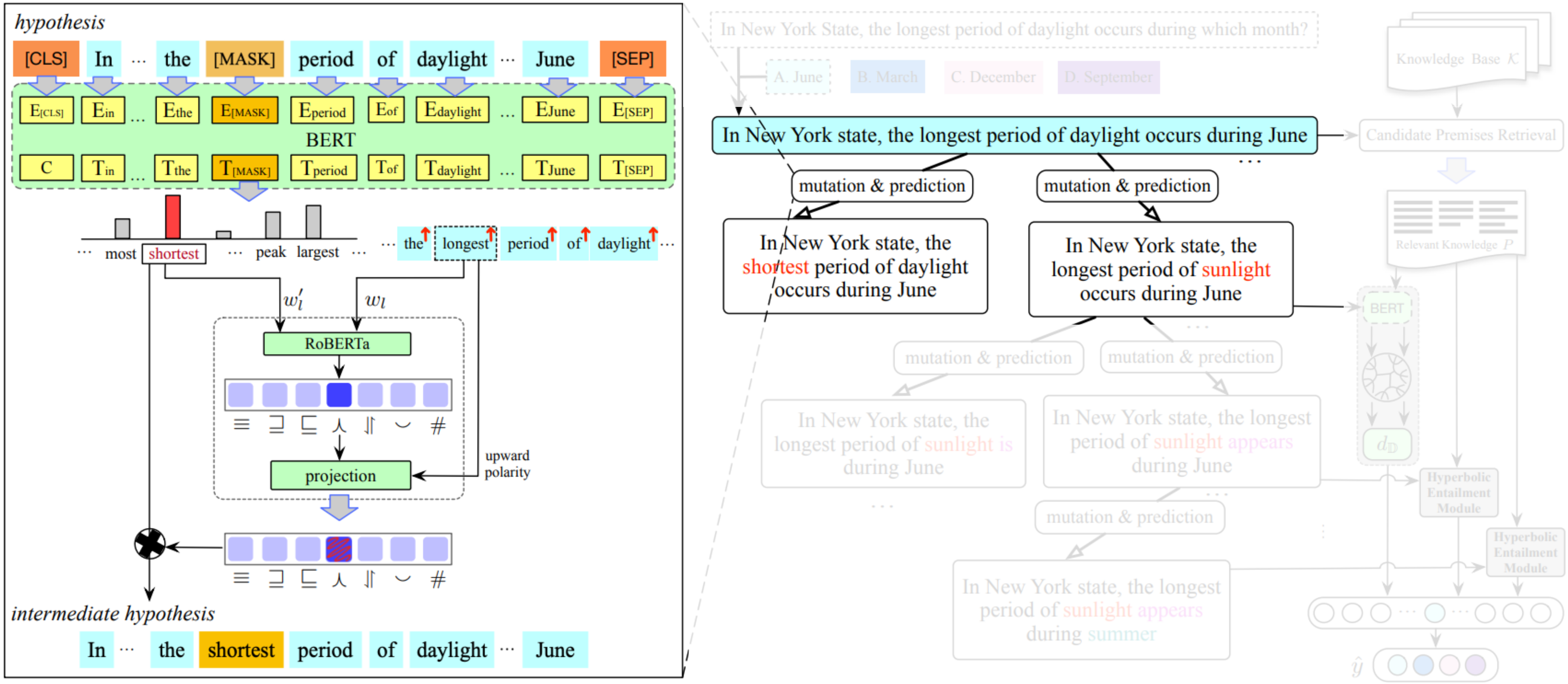


Candidate Premises Retrieval

1. representation of h_i and each p_j in KB by computing the average Glove word embeddings
2. calculate the cosine similarity between h_i and each p_j to find the top k relevant candidate premises

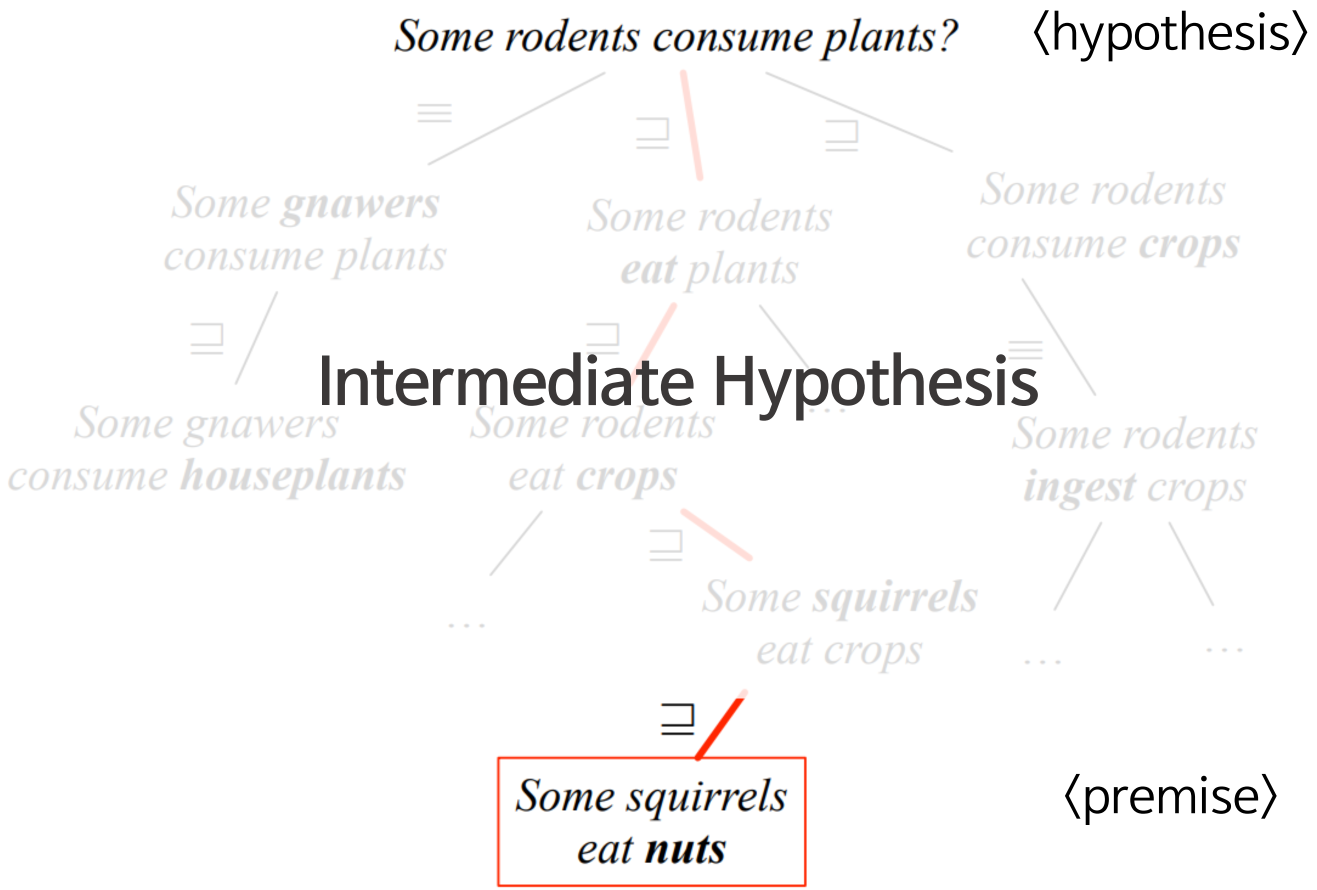


Candidate Proof Path Generation



Methods : Candidate Proof Path Generation

Candidate Proof Path Generation



$$\frac{\begin{array}{c}\text{Intermediate Hypothesis} \sqsupseteq \text{Hypothesis} \\ \text{Premise} \sqsupseteq \text{Intermediate Hypothesis}\end{array}}{\text{Premise} \sqsupseteq \text{Hypothesis}}$$

Methods : Candidate Proof Path Generation

Filter out little influence on the semantics of the hypothesis

NLTK (Bird et al., 2009) toolkit

preposition, determiner, coordinating conjunction, cardinal numbers, personal pronoun, modal verb, punctuation words, stop words

Methods : Candidate Proof Path Generation

Mutation for Intermediate Hypothesis

1. Substitution

Hypothesis	In New York state, the longest period of daylight occurs during June
Masked Hypothesis	In New York state, the [MASK] period of daylight occurs during June

2. Insertion/Deletion

Adjective을 inset or delete

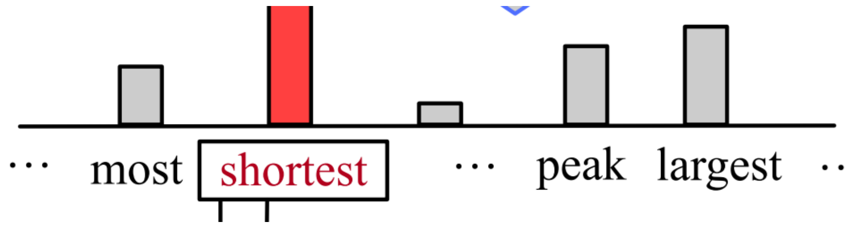
Noun 앞에 [mask]

Hypothesis	In New York state, the longest period of daylight occurs during June
Masked Hypothesis	In [MASK] New York state, the longest period of daylight occurs during June

Mutation for Intermediate Hypothesis

Masked Hypothesis	In New York state, the [MASK] period of daylight occurs during June

BERT



Candidate Words for [MASK]

Methods : Proof Path Filtering

Proof Path Filtering

Mask mechanism does not guarantee semantic coherence

Hypothesis	In New York state, the longest period of daylight occurs during June
Highest Prob	In New York state, the shortest period of daylight occurs during June

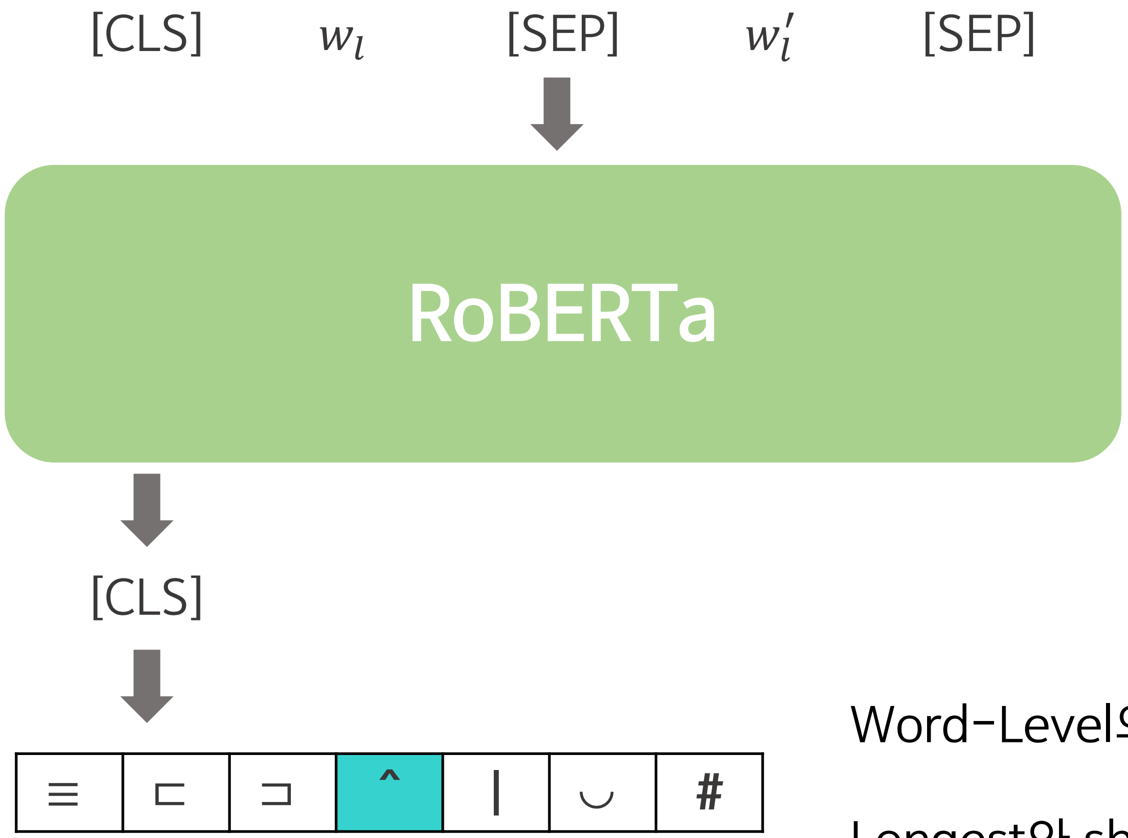
→ Substitution, Insertion, Deletion 문맥적으로 뜻이 달라질 수 있다.

우리는 비슷한 Semantic으로 추론해가고 있다.

Original hypothesis의 semantic을 지켜야 한다!

Methods : Proof Path Filtering

Original word	w_l : longest
Candidate word	w_l' : shortest



Word-Level의 relation을 classification

Longest와 shortest의 관계는 negation

Methods : Proof Path Filtering

Semantic을 유지하는 가를 보기 위해 sentence-level 확인

Original word가 positive polarity면 sentence level에서 semantic이 변하지 않을 것이다.

Original word의 ↓polarity와 predicted relation으로 결정

Polarity tagging은 stanford의 natlog 사용

실제 inference에 사용하는 relation

r	$\equiv$	$\sqsubset$	$\sqsupset$	$\wedge$	$\mid$	$\smile$	$\#$
$\phi(r)$	$\equiv$	$\sqsupset$	$\sqsubset$	$\mid$	$\wedge$	$\smile$	$\#$

↓로 relation이 바뀌는 경우

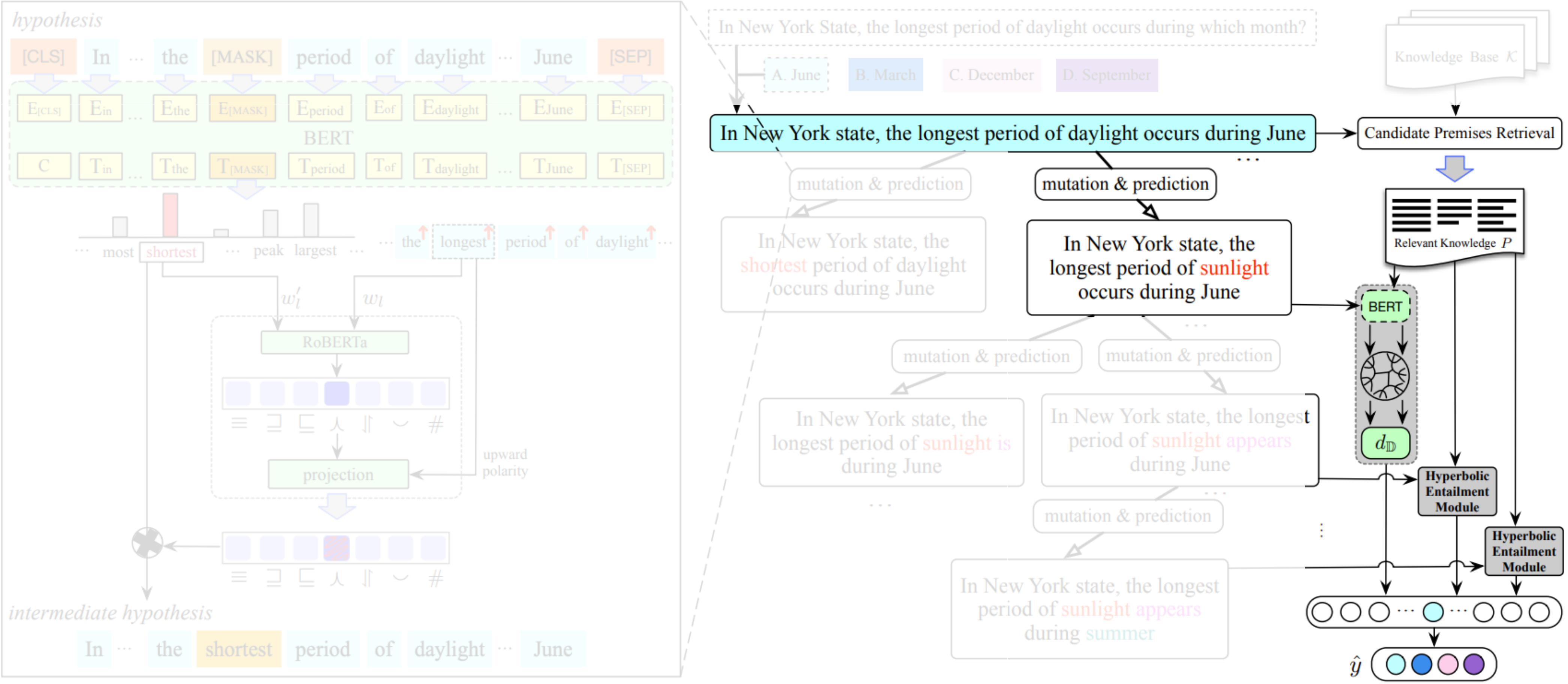
≡	⊏	⊐	^	|	‿	#

〈hypothesis〉

the ↑longest period of daylight occurs during June

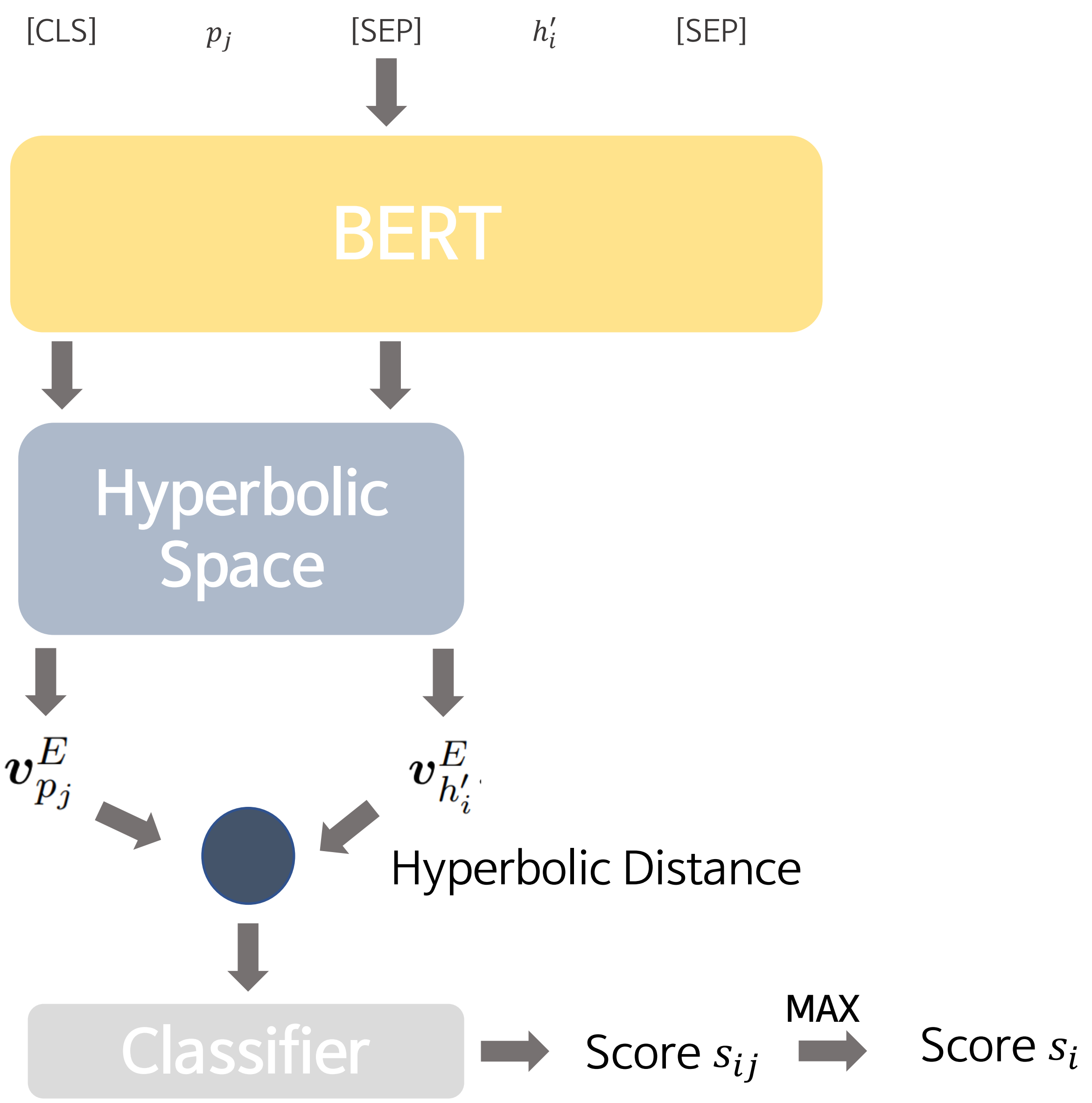
Methods : Candidate Proof Path Generation

Candidate Proof Path Generation



Methods : Entailment Score Estimation in Hyperbolic Space

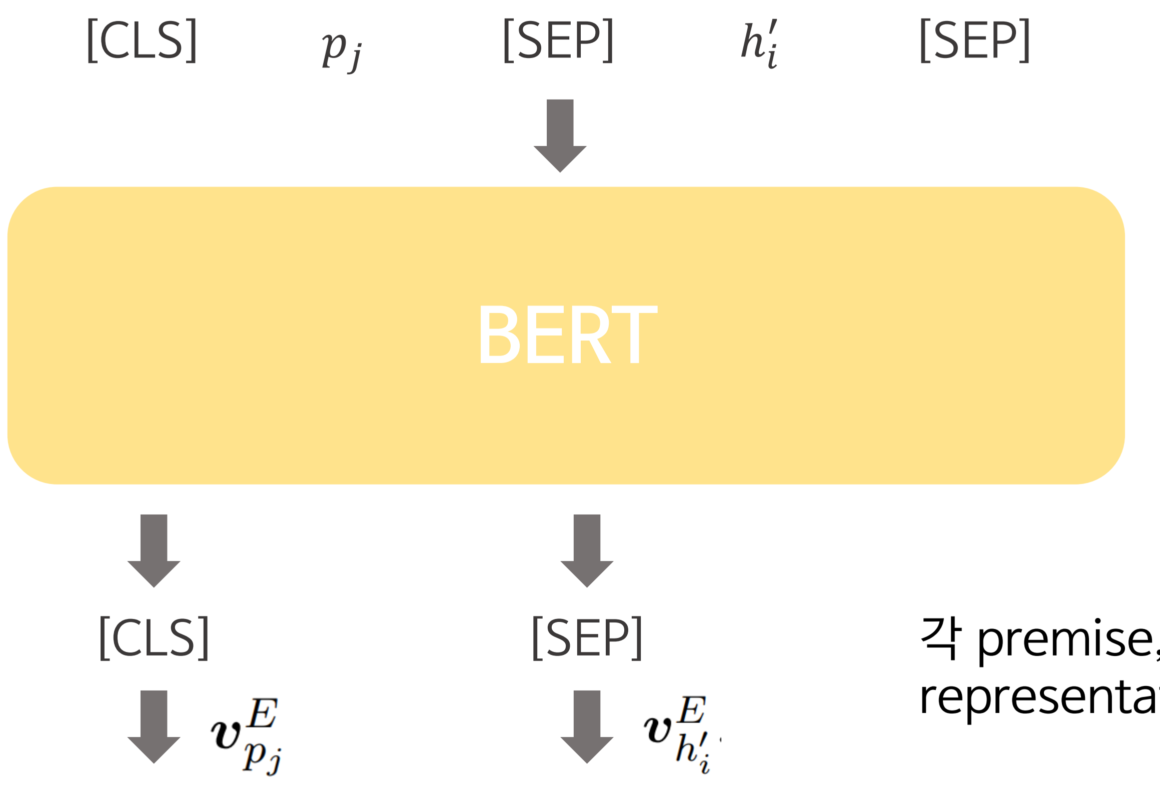
Candidate Premise	p_j
Intermediate Hypothesis	h'_i



Entailment Score Estimation in Hyperbolic Space

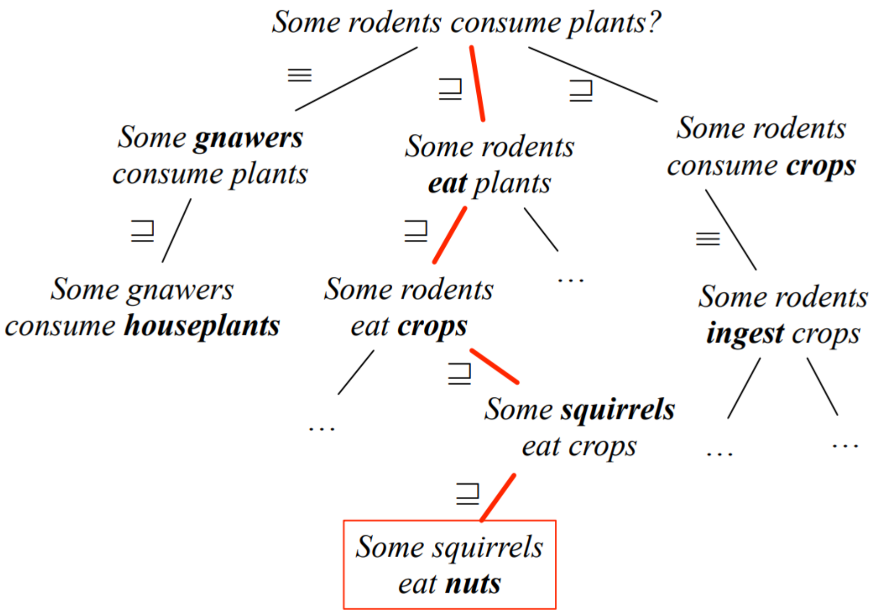
Intermediate hypothesis와 candidate premise간 score 측정

Candidate Premise	p_j
Intermediate Hypothesis	h'_i



각 premise, intermediate hypothesis의 representation의 거리로 score를 구하면 된다!

Intermediate Hypothesis grows **exponentially** but Euclidean space grows **polynomially**



Methods : Entailment Score Estimation in Hyperbolic Space

Intermediate Hypothesis grows **exponentially** but Euclidean space grows **polynomially**

$$
\begin{aligned}
\overline{\mathbf{m}}_{p_j} &= \psi_{\text{dir}}\left(\boldsymbol{v}^{E}_{p_j}\right), & \mathbf{m}_{p_j} &= \frac{\overline{\mathbf{m}}_{p_j}}{\left\|\overline{\mathbf{m}}_{p_j}\right\|} \\
\bar{\mu}_{p_j} &= \psi_{\text{norm}}\left(\boldsymbol{v}^{E}_{p_j}\right), & \mu_{p_j} &= \sigma\left(\bar{\mu}_{p_j}\right)
\end{aligned}
$$

$\psi_{dir} : \mathbb{R}^{d} \rightarrow \mathbb{R}^{d_H}$ is a multi-layer

$\psi_{norm} : \mathbb{R}^{a} \rightarrow \mathbb{R}$ is a linear function

$$
\boldsymbol{v}^{H}_{p_j} = \mu_{p_j}\mathbf{m}_{p_j}
$$

Methods : Entailment Score Estimation in Hyperbolic Space

Hyperbolic distance

$$d_{\mathbb{D}}(\boldsymbol{v}^H_{p_j}, \boldsymbol{v}^H_{h'_i}) = \cosh^{-1}\left(1 + 2\frac{\|\boldsymbol{v}^H_{p_j} - \boldsymbol{v}^H_{h'_i}\|^2}{(1 - \|\boldsymbol{v}^H_{p_j}\|^2)(1 - \|\boldsymbol{v}^H_{h'_i}\|^2)}\right)$$

vector

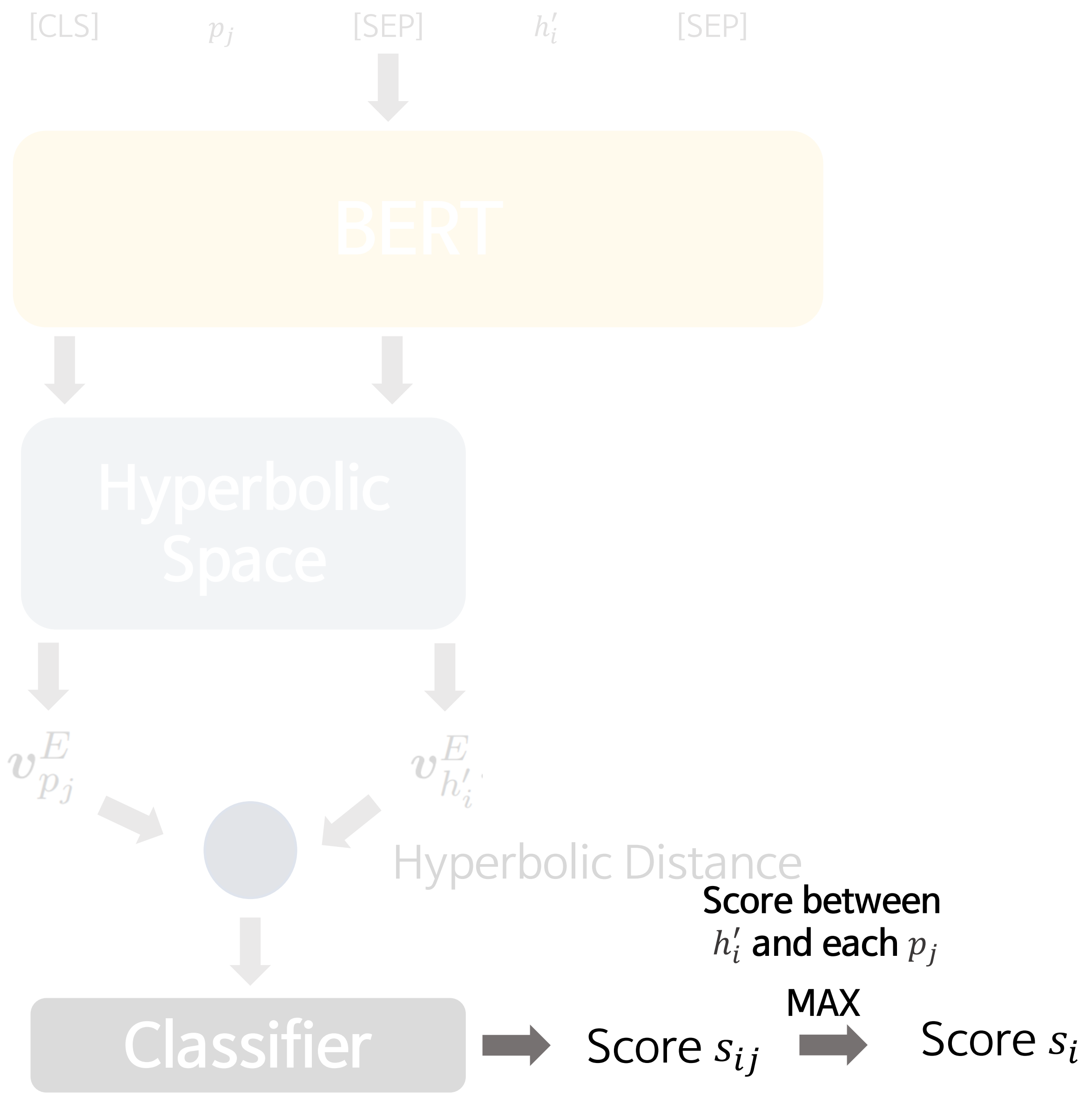
→ CLASSIFIER → Scalar (Score)

```
# entailmentscore
self.hyperdis_prob_classify = nn.Linear(1, 1)
output = self.poincare_distance(CLS_hyperbolic_vector, SEP_hyperbolic_vector)

sim_score = self.hyperdis_prob_classify(output.unsqueeze(dim=1))
```

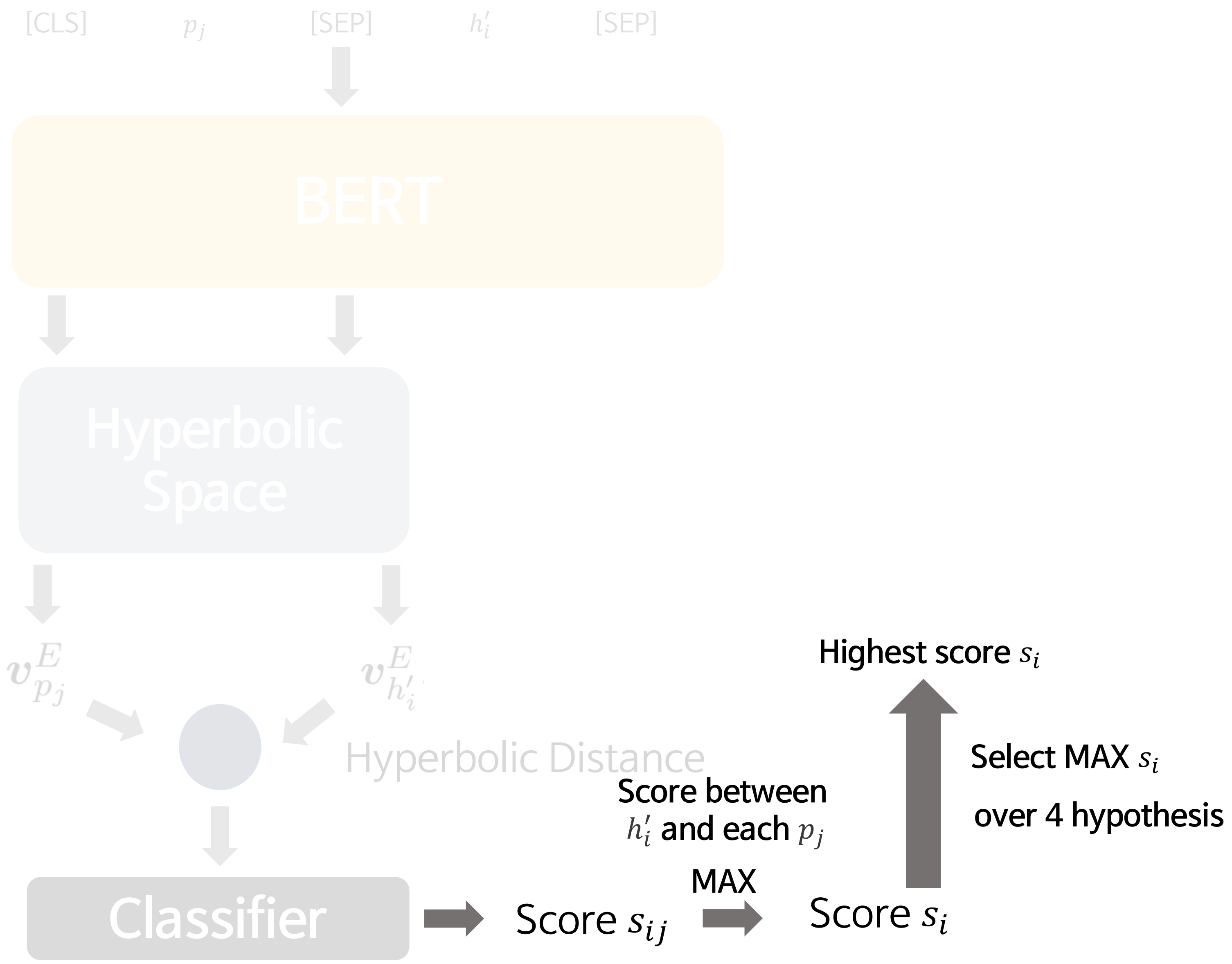
Methods : Entailment Score Estimation in Hyperbolic Space

Candidate Premise	p_j
Intermediate Hypothesis	h'_i



Methods : Entailment Score Estimation in Hyperbolic Space

Candidate Premise	p_j
Intermediate Hypothesis	h'_i



Experiment

Dataset

: QA-S & QA-L

multiple-choice science questions

from the New York Regents 4th Grade Science Exams

	QA-S	**QA-L**
Train	108	500
Val	61	249
Test	68	250

knowledge bases

: Barron's and SCITEXT

Experiment

QA-S

Accuracy (%)

Model	Barron's	SCITEXT
Solr Only	42	58
Classifier	52	60
+ Solr	48	64
Evaluation Function	54	63
+ Solr	45	58
NaturalLI (Angeli et al., 2016)	51	61
+ Solr	49	61
+ Solr + Classifier	49	67
HyperQA (Tay et al., 2018)	54	62
SemBERT (Zhang et al., 2020)	53	59
NeuNLI-E (Ours)	57	67
NeuNLI (Ours)	**64***	**72***

QA-L

Model	Accuracy
Solr Only	46.8
Classifier	43.6
NaturalLI (Angeli et al., 2016)	46.4
+ Solr	48.0
HyperQA (Tay et al., 2018)	47.6
SemBERT (Zhang et al., 2020)	47.2
NeuNLI-E (Ours)	48.8
NeuNLI (Ours)	**50.8***

Experiment

Relation Prediction between masked words and original word

BERT보다 RoBERTa가 더 좋았다.

Relation	**BERT**			**RoBERTa**		
	P	R	F1	P	R	F1
equivalence	0.79	0.83	0.81	0.81	0.85	0.83
forward entailment	0.75	0.69	0.72	0.75	0.74	0.75
reverse entailment	0.69	0.69	0.69	0.70	0.72	0.71
negation	0.74	0.58	0.65	0.85	0.63	0.72
alternation	0.54	0.64	0.58	0.57	0.61	0.59
cover	0.42	0.32	0.36	0.48	0.32	0.39
independence	0.63	0.58	0.60	0.66	0.62	0.64

Table 6: Performance of lexical relation prediction.

Experiment

Human Evaluation for Explainability

on the QA-S dataset with the Barron's knowledge base

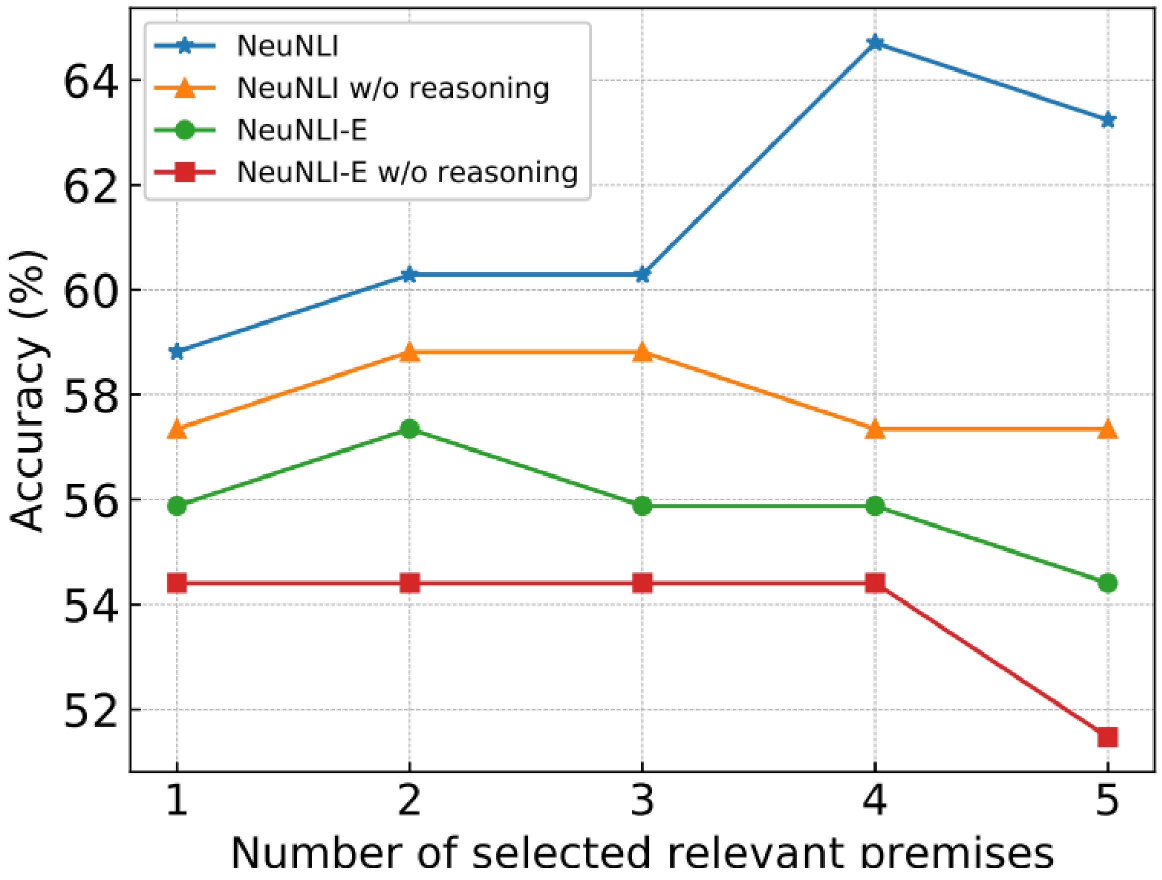
3명의 NLP 전공 졸업생 {0, 1, 2} inference path에 점수를 주었다.

Path가 설명을 잘했다면 2 아니라면 0 그 사이는 1

Baseline(NaturalLI)와 our model(NeuNLI)의 inference path 점수 비교

	NaturalLI	**NeuNLI**
Avg. Explainability Score	1.09	1.31*

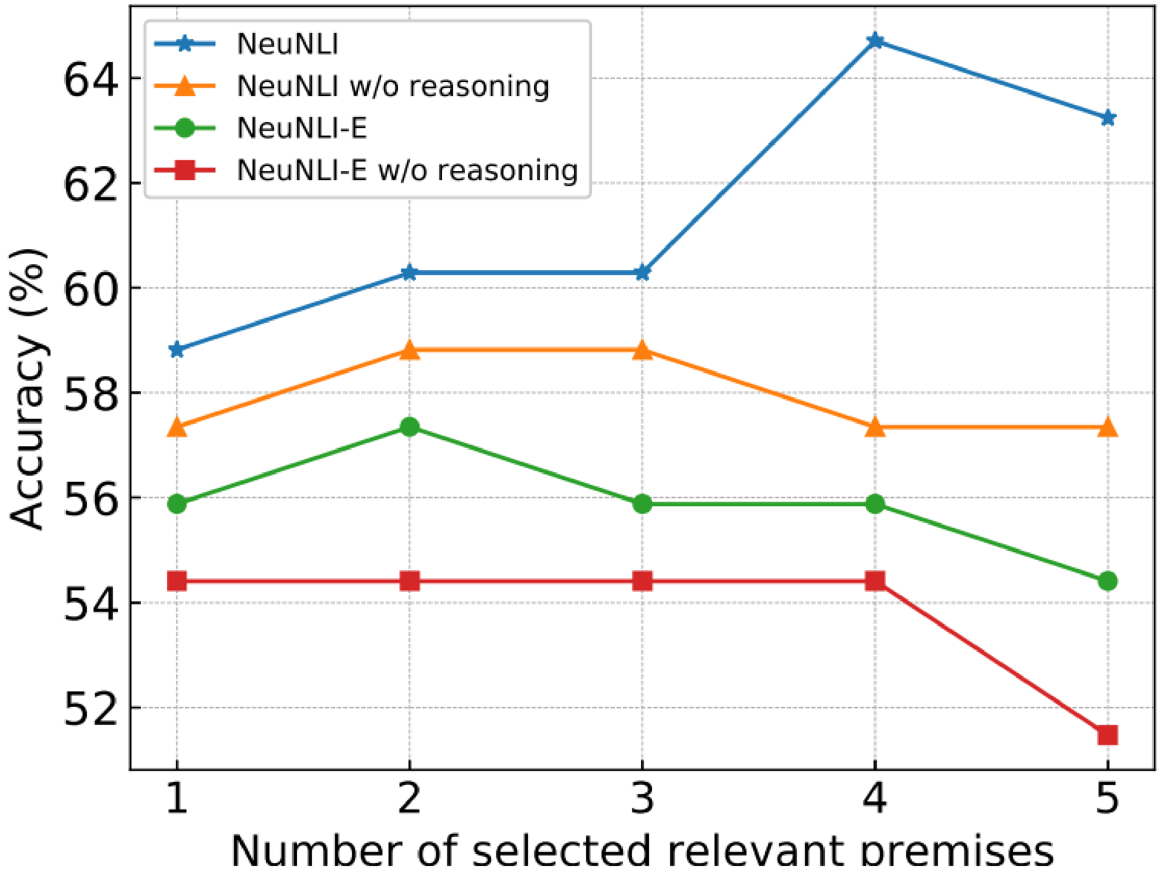
(1) Effect of Number of Relevant



Experiment

(2) Effectiveness of Natural Logic-based Reasoning

Inference path의 효과 확인



Summary

Natural Logic을 사용한 Open Domain QA

1. Retrieve Relevant Premise

    K premises, Corresponding hypothesis

2. Get Intermediate Hypothesis

    by using natural logic

3. Calculate Semantic Score

    between intermediate hypothesis and each premise

Critiques

Strength

- 전통적인 추론 풀이 방식인 Natural Logic을 Neural Net을 쓰면서 적용함
- Natural Logic을 이용하여 모델이 어떤 논리로 추론을 하는지 쉽게 알 수 있다.

Weakness

- Baseline benchmark외 benchmark 성능 확인 없음
- 사용하지 않는 예시를 들어 논문으로 보았을 때 실제로 그 explanability를 확인하기 힘듦

Question?

Thank You!