

BERT 사전 훈련 및 미세 조정

MediaZen

2022년 9월 17일 토요일

CONTENTS

01 사전 훈련(Pre-training)

02 미세 조정(Fine-tuning)

03 정리



사전 훈련 (Pre-training)

1) 마스크드 언어 모델

2) 다음 문장 예측

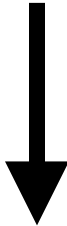
1) 마스크드 언어 모델(Masked Language Model, MLM)

나는 [MASK]에 가서 그곳에서 빵과 [MASK]를 샀다.

1) 마스크드 언어 모델(Masked Language Model, MLM)

- 80%의 단어들은 [MASK]로 변경한다.
ex) The man went to the store → The man went to the [MASK]
- 10%의 단어들은 랜덤으로 단어가 변경된다.
ex) The man went to the store → The man went to the dog
- 10%의 단어들은 동일하게 둔다.
ex) The man went to the store → The man went to the store

My dog is cute. He likes playing.



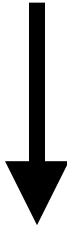
토큰화

my dog is cute he likes play ##ing

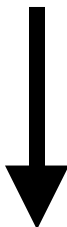


마스킹

my [MASK] is cute king likes play ##ing



BERT



my ? is cute ? likes ? ##ing

2) 다음 문장 예측(Next Sentence Prediction, NSP)

- 이어지는 문장의 경우

ex) Sentence A : The man went to the store.

Sentence B : He bought a gallon of milk.

Label = **IsNextSentence**

- 이어지는 문장이 아닌 경우

ex) Sentence A : The man went to the store.

Sentence B : The dogs are so cute.

Label = **NotNextSentence**

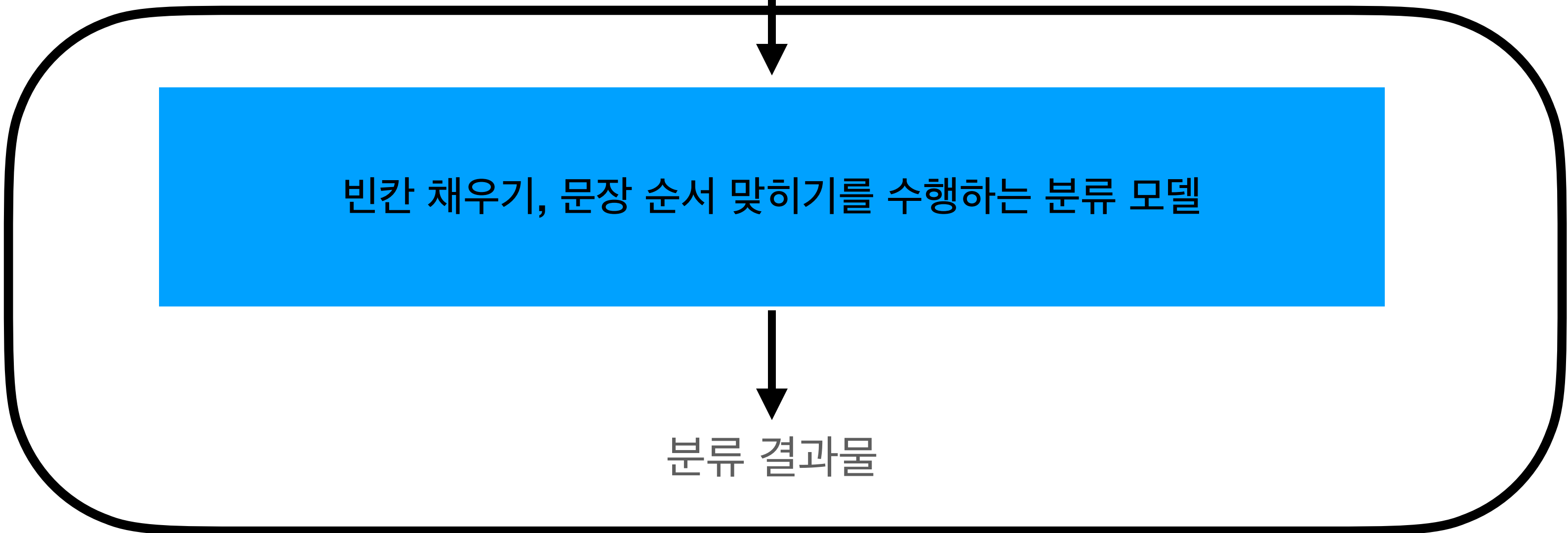
"지난" "8" "월" "시작" "된" ... (생략)



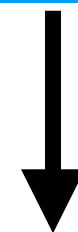
인코딩 (임베딩) : 각 토큰을 벡터로 만듦



"지난"을 나타내는 벡터 "8"을 나타내는 벡터 "월"을 나타내는 벡터 "시작"을 나타내는 벡터 "된"을 나타내는 벡터 ... (생략)



빈칸 채우기, 문장 순서 맞히기를 수행하는 분류 모델



분류 결과물

프리트레인 학습이
완료되면 버림

미세 조정 (Fine-tuning)

1) 의도 분류

2) 질의 응답

1) 의도 분류
(Intent Classification)

[CLS] 아이유 노래 아무거나 틀어줘 [SEP]

[CLS] 를
나타내는 벡터

1	0	1	2
---	---	---	---

(토큰 수) X 4 행렬

X

0.5	0.5	0
0.2	0	0
0.5	0.5	0.2
0.5	0	0.5

4 X (의도 수) 행렬
fine-tuning layer

=

[CLS]

가수이름 으로 노래	노래제목 으로 노래	노래중지
---------------	---------------	------

2	1	1.2
---	---	-----

(토큰 수) X (의도 수) 행렬

1) 의도 분류
(Intent Classification)

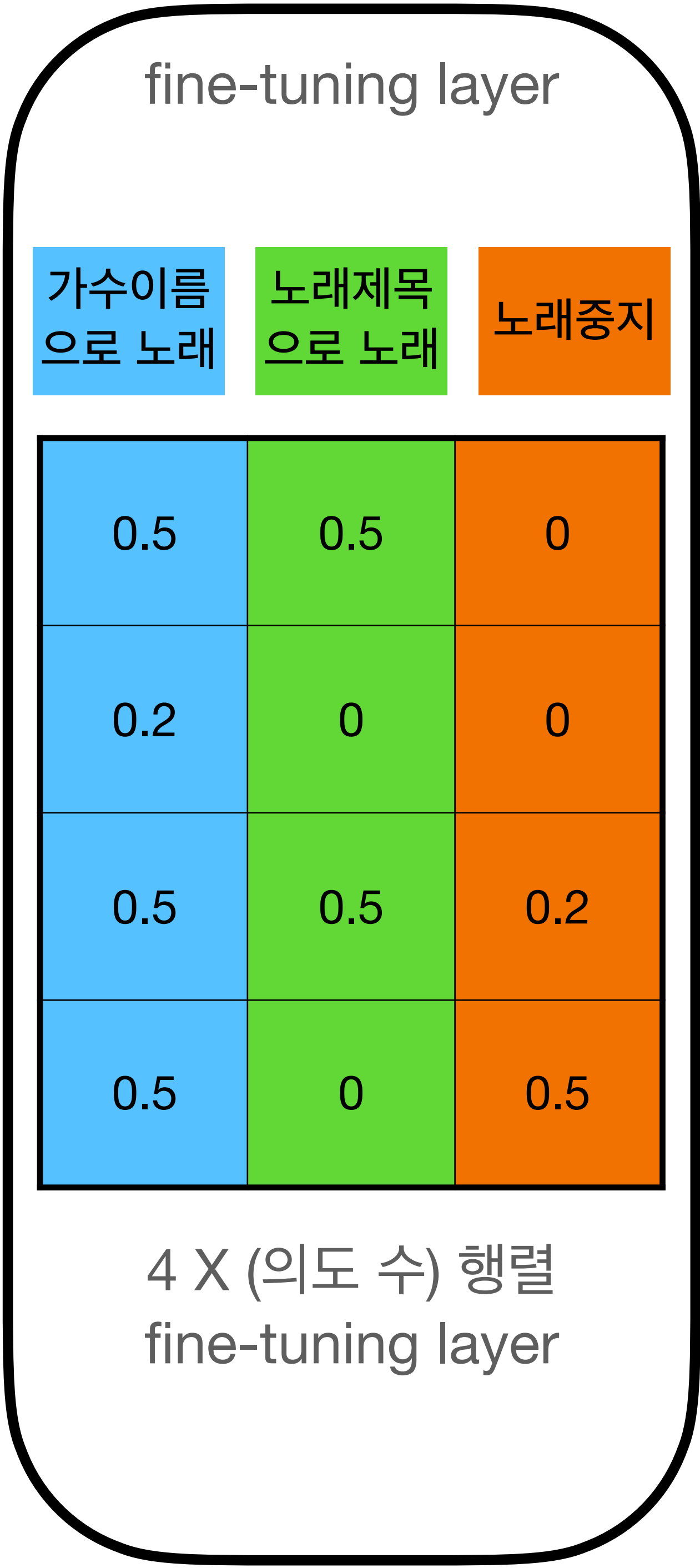
[CLS] 아이유 노래 아무거나 틀어줘 [SEP]

[CLS] 를
나타내는 벡터

1	0	1	2
---	---	---	---

(토큰 수) X 4 행렬

X



=

[CLS]

가수이름
으로 노래

노래제목
으로 노래

노래중지

2	1	1.2
---	---	-----

(토큰 수) X (의도 수) 행렬

[CLS] 아이유 노래 아무거나 틀어줘 [SEP]

[CLS]	가수이름 으로 노래	노래제목 으로 노래	노래중지
	2	1	1.2

(토큰 수) X (의도 수) 행렬

softmax

[CLS]	가수이름 으로 노래	노래제목 으로 노래	노래중지
	0.55↑	0.2↓	0.25↓

(토큰 수) X (의도 수) 행렬

정답	1	0	0
----	---	---	---

모델의 최종 아웃풋

1) 의도 분류
(Intent Classification)

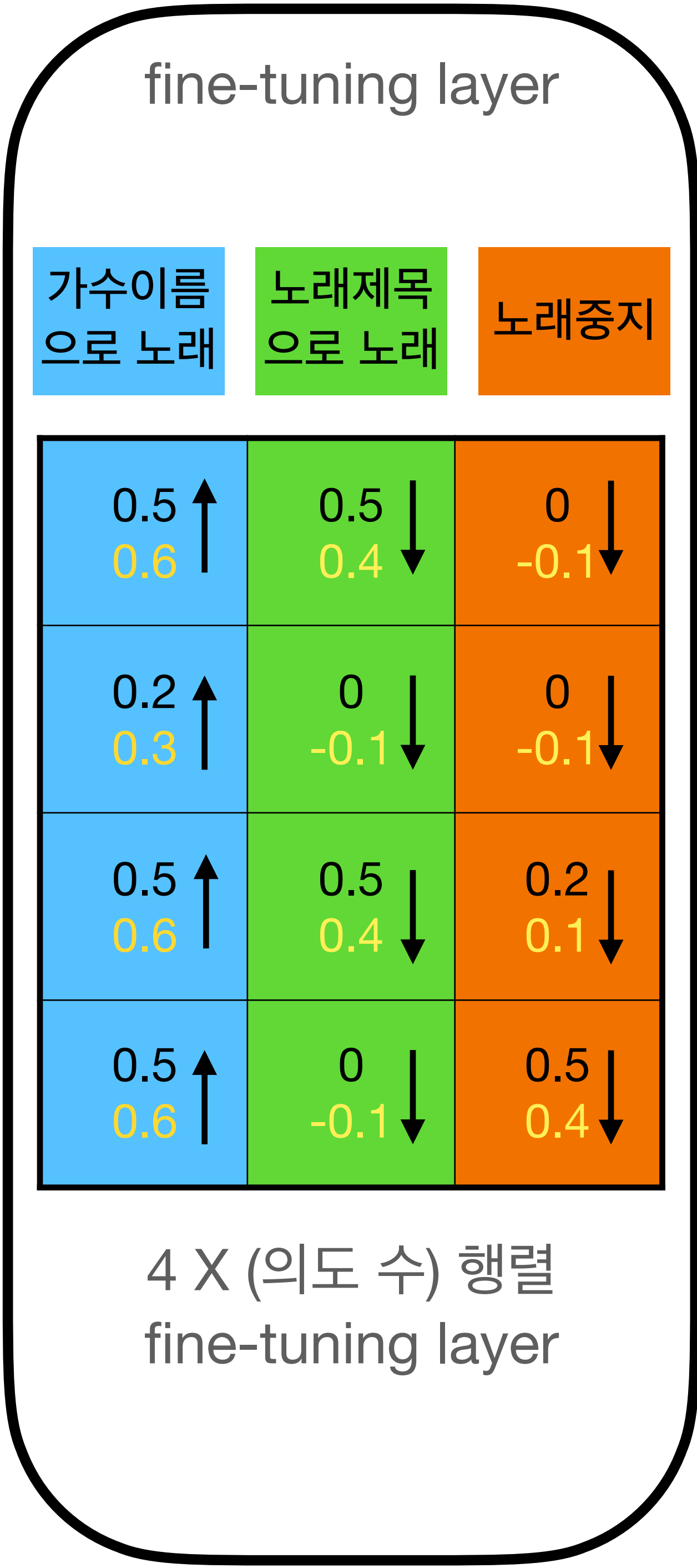
[CLS] 아이유 노래 아무거나 틀어줘 [SEP]

[CLS] 를 나타내는 벡터

1	0	1	2
---	---	---	---

(토큰 수) X 4 행렬

X



=

[CLS]

가수이름
으로 노래 노래제목
으로 노래 노래중지

2 2.4	1 0.6	1.2 0.8
----------	----------	------------

(토큰 수) X (의도 수) 행렬

1) 의도 분류
(Intent Classification)

[CLS] 노래 멈춰줄래 [SEP]

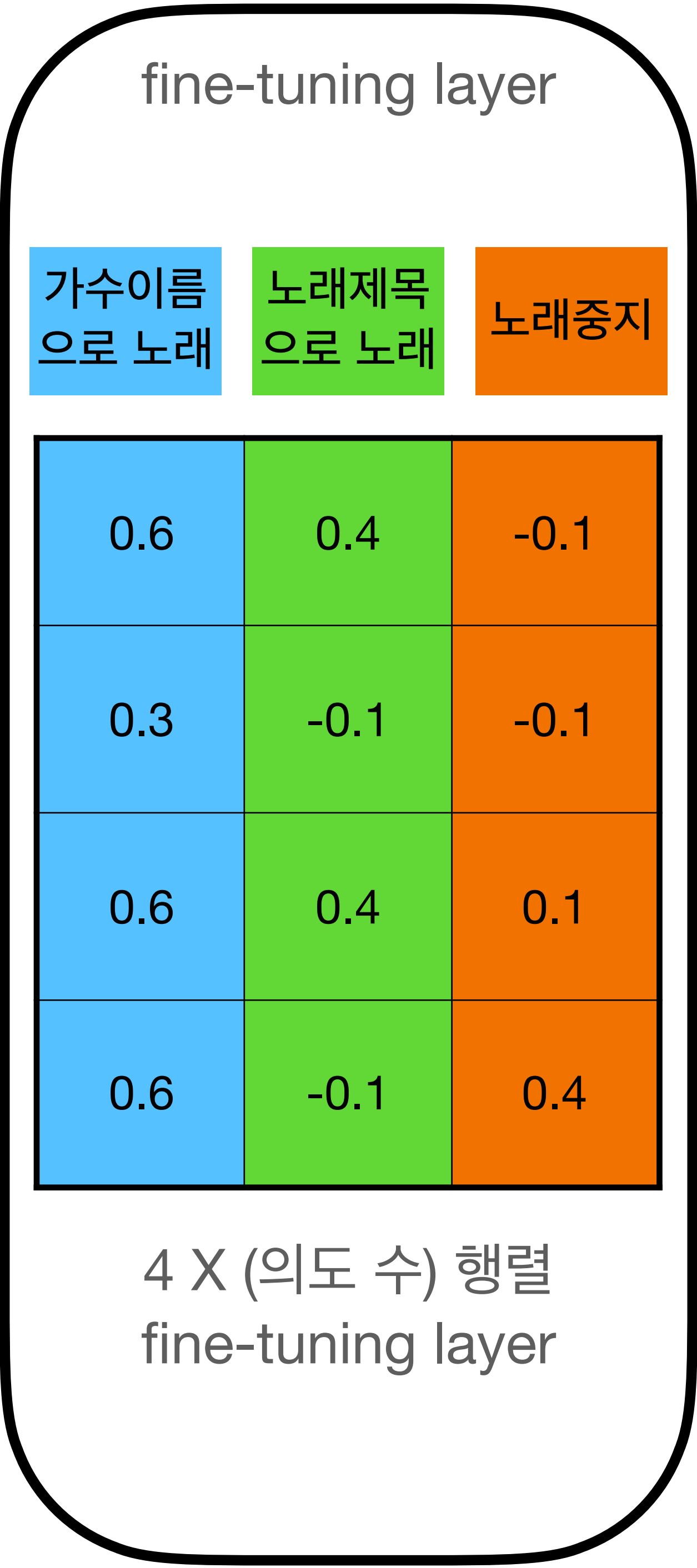


[CLS] 를
나타내는 벡터

1	1	0	1
---	---	---	---

(토큰 수) X 4 행렬

X



=

[CLS]

가수이름
으로 노래

노래제목
으로 노래

노래중지

1.5	0.2	0.2
-----	-----	-----

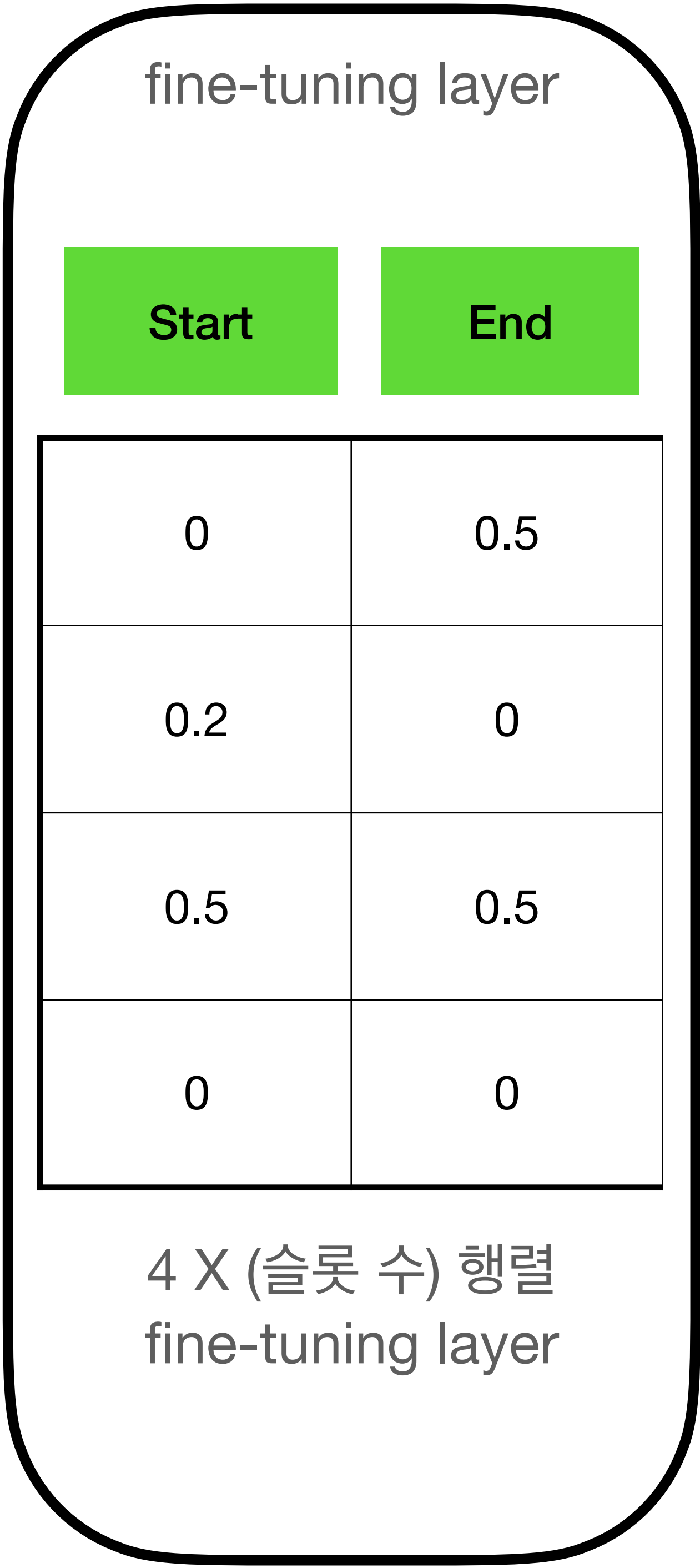
(토큰 수) X (의도 수) 행렬

2) 질의 응답(QA)

“이순신”를 나타내는 벡터	1	0	0	0.5
“은”를 나타내는 벡터	0	1	1	1
“1545”을 나타내는 벡터	1	2	2	1
“년”를 나타내는 벡터	2	0	2	0
“태어나”을 나타내는 벡터	0	1	2	2
“...”

(토큰 수) X 4 행렬

X



=

	Start	End
“이순신”	0	0.5
“은”	0.7	0.5
“1545”	1.4	1.5
“년”	1	2
“태어나”	1.2	1
“...”

(토큰 수) X (슬롯 수) 행렬

	Start	End
“이순신”	0	0.5
“은”	0.7	0.5
“1545”	1.4	1.5
“년”	1	2
“태어나”	1.2	1
“ ... ”

(토큰 수) X (슬롯 수) 행렬

softmax →

모델의 최종 아웃풋		
	Start	End
“이순신”	0.1	0.1
“은”	0.15	0.1
“1545”	0.3	0.25
“년”	0.2	0.4
“태어나”	0.25	0.15
“ ... ”

(토큰 수) X (슬롯 수) 행렬

정리

한눈에 살펴보기

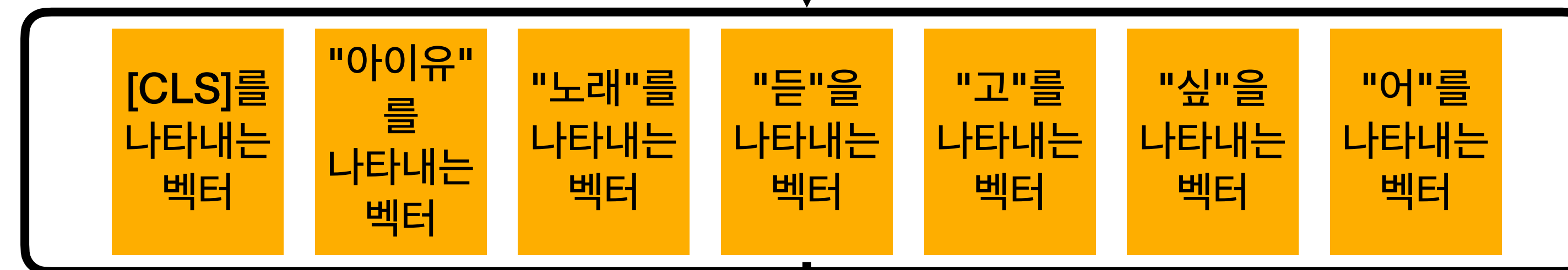
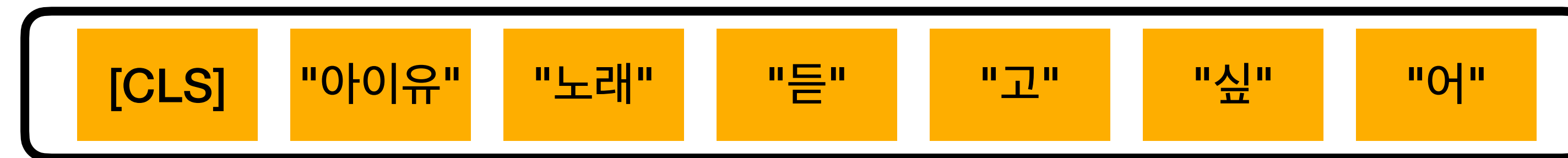
- **사전 훈련(Pre-training)**

- 1) 마스크드 언어 모델 나는 [MASK]에 가서 그곳에서 빵과 [MASK]를 샀다.
- 2) 다음 문장 예측하기 Sent A, Sent B, IsNextSentence

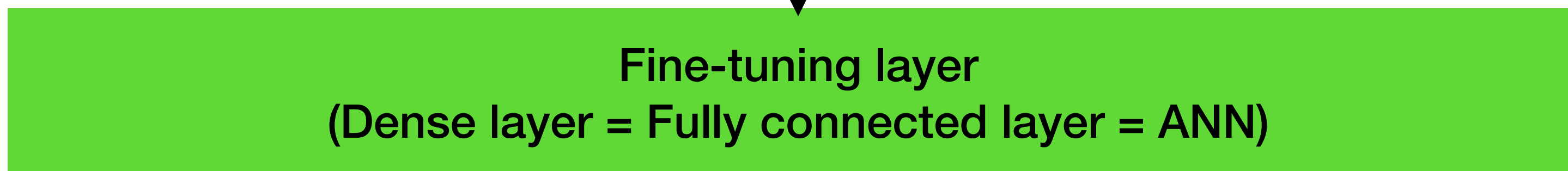
- **미세 조정(Fine-tuning)**

- 1) 슬롯 태깅 가수 이름, 노래 제목, 슬롯 X
- 2) 의도 분류 이름으로 노래 요청, 제목으로 노래 요청, 노래 중지
- 3) 질의 응답 이순신 언제 태어났나요? 1545년

슬롯 태깅
& 의도 분류



Pre-trained 모델의 출력이면서
Fine-tuning layer의 입력

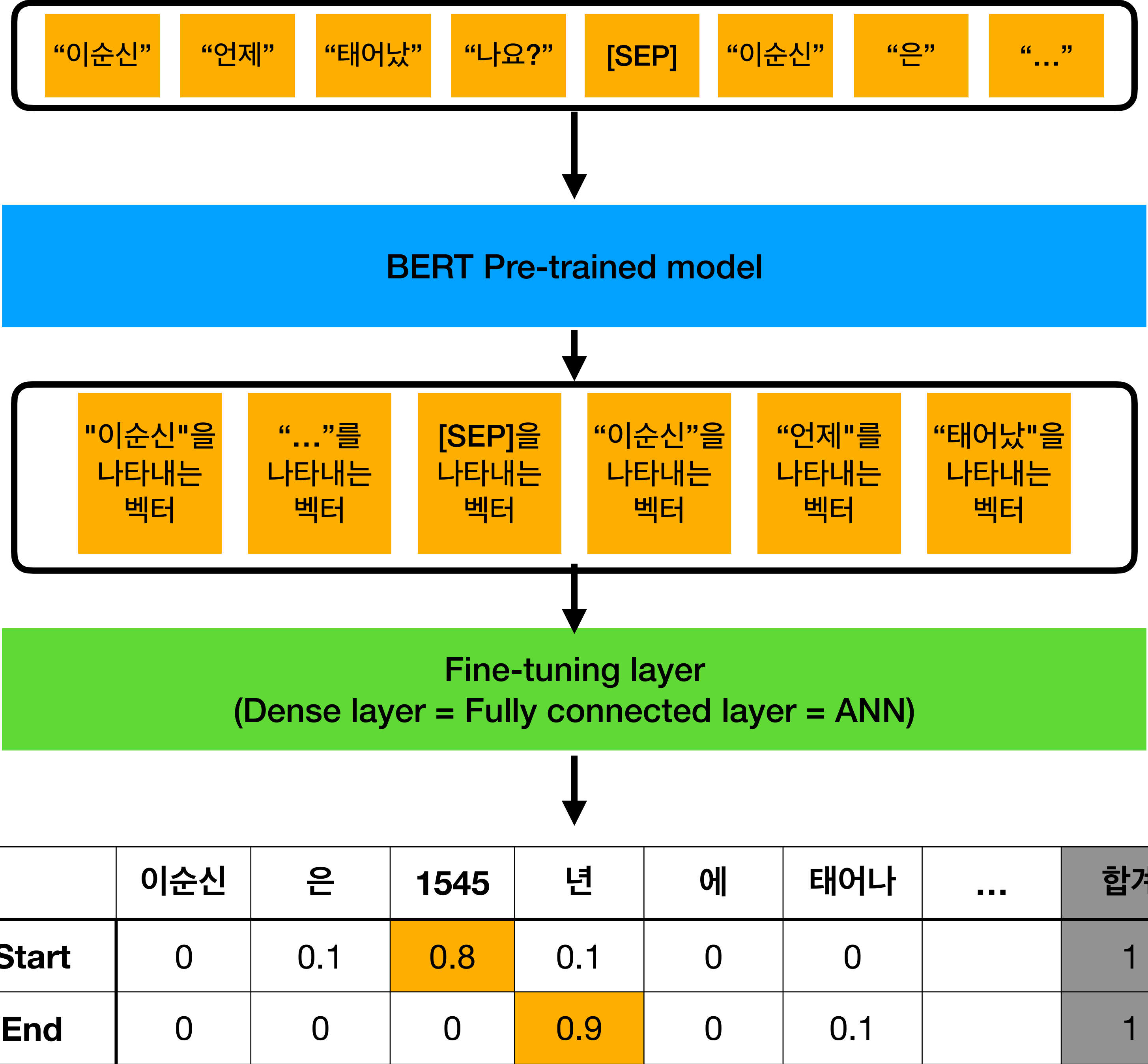


	[CLS]
가수 이름으로 노래 요청	0.9
노래 제목으로 노래 요청	0.1
노래 중지 요청	0
합계	1

	아이유	노래	듣	고	싶	어
가수 이름	0.8	0.1	0	0	0	0
노래 제목	0.2	0.4	0.4	0.2	0.4	0.4
중요 슬롯 아님	0	0.5	0.6	0.8	0.6	0.6
합계	1	1	1	1	1	1

Fine-tuning layer의 출력

질의 응답



Pre-trained 모델의 출력이면서
Fine-tuning layer의 입력

Fine-tuning layer의 출력

수고 많았습니다.