AI VIET NAM
@aivietnam.edu.vn

Module 08 – Project

Aspect-based Sentiment Analysis

Code - Data

Nguyen Quoc Thai
MSc in Computer Science

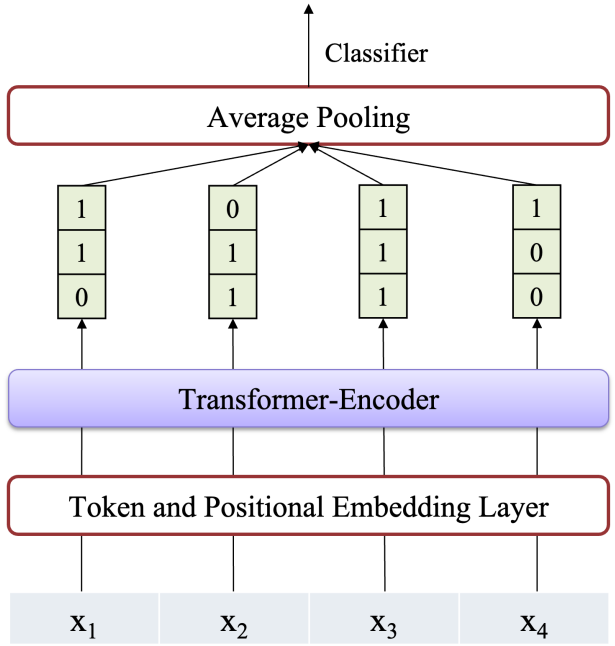
Objectives

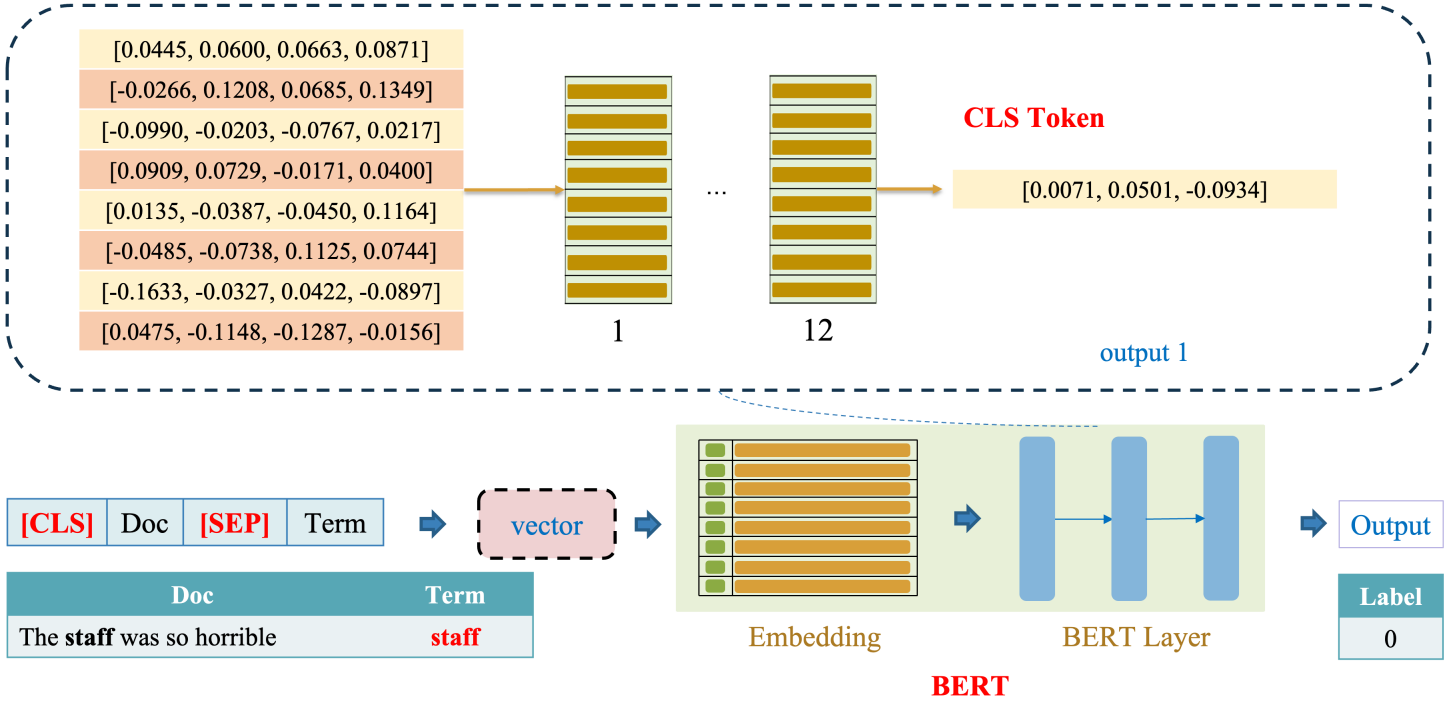
Sentiment Analysis

- Text Classification
- Types of Sentiment Analysis
- Classifier Model

Aspect-based Sentiment Analysis

- Subtasks
- Aspect Term Extraction
- Aspect Sentiment Pair Extraction





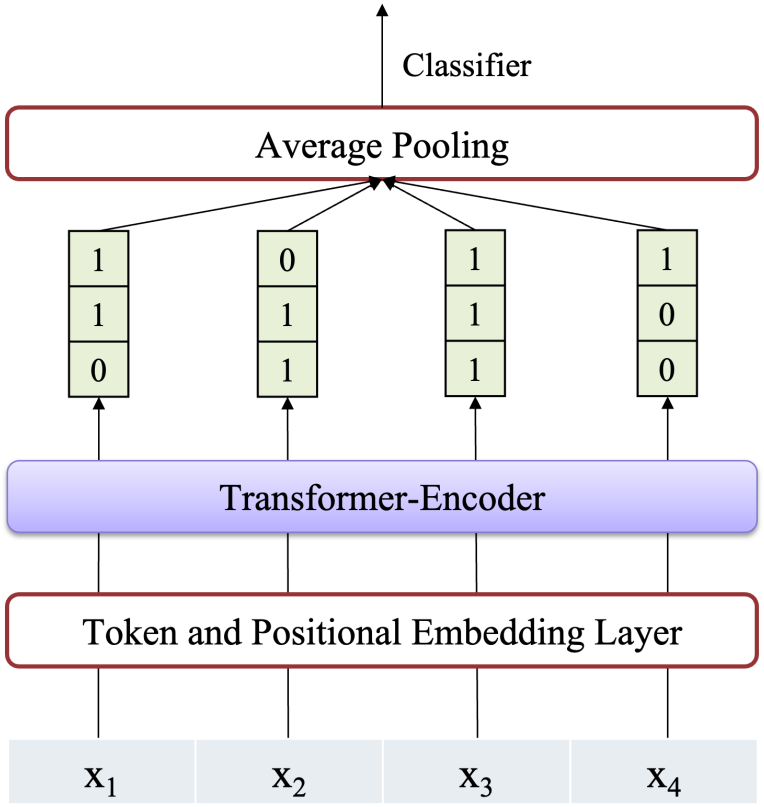
Outline

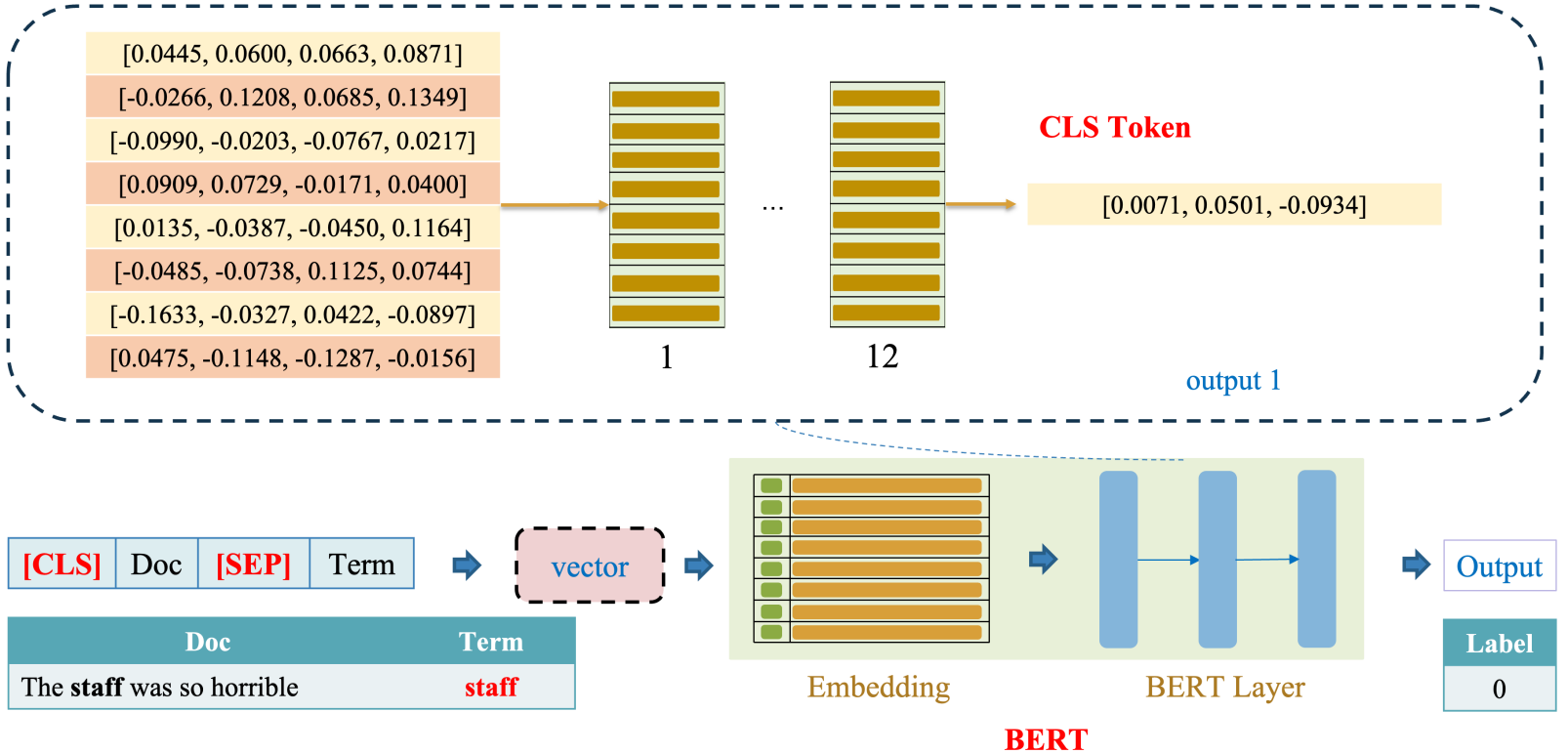
SECTION 1

Sentiment Analysis

SECTION 2

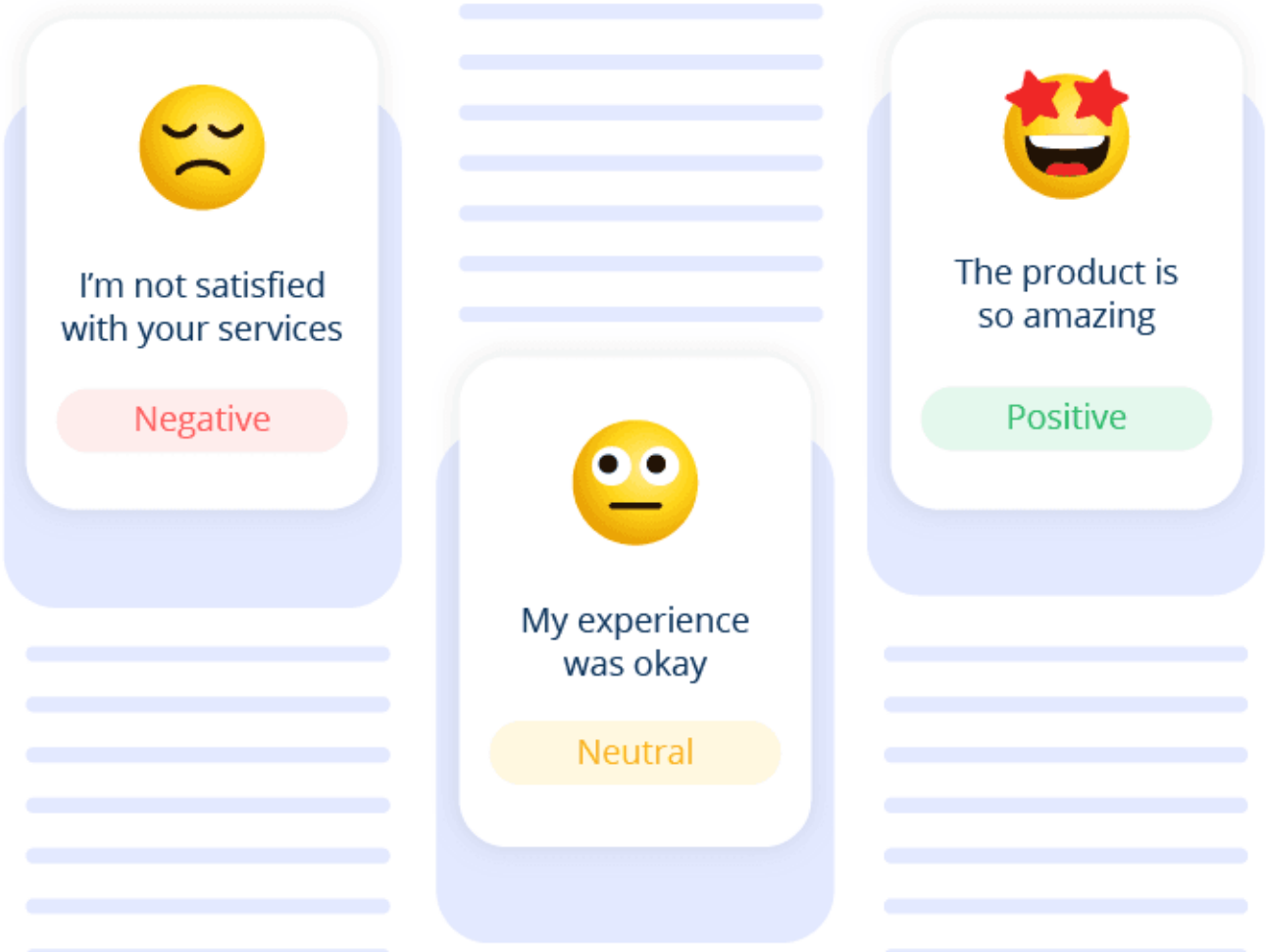
Aspect-based SA





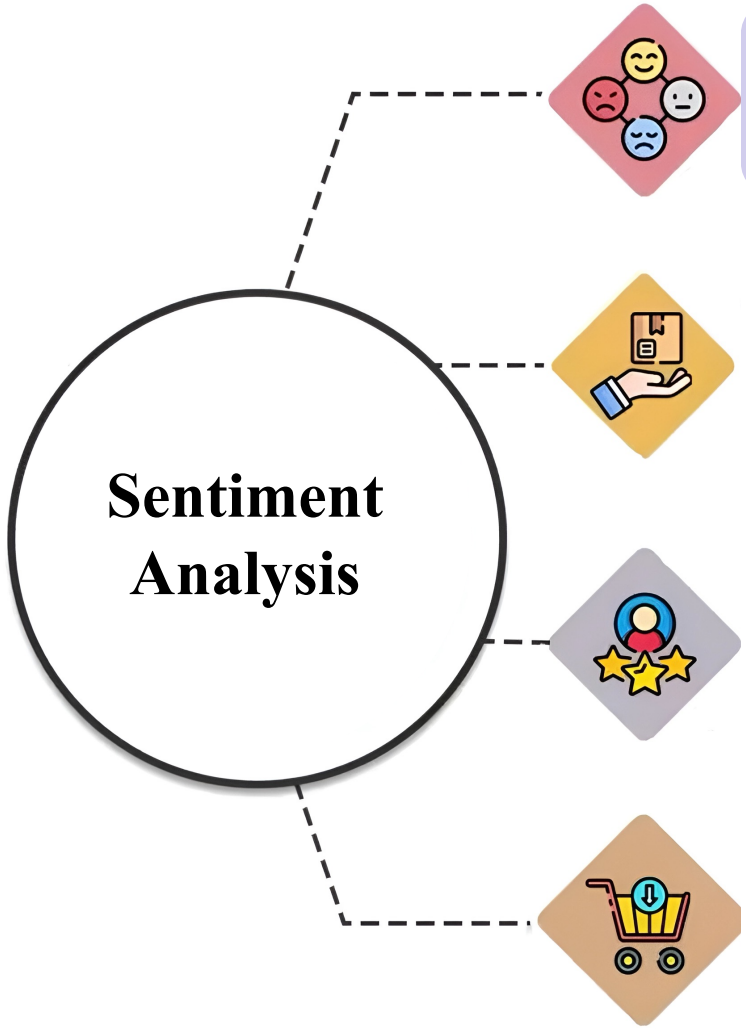
Sentiment Analysis

Text Classification



Sentiment Analysis

Types of Sentiment Analysis



Emotion Detection Sentiment Analysis

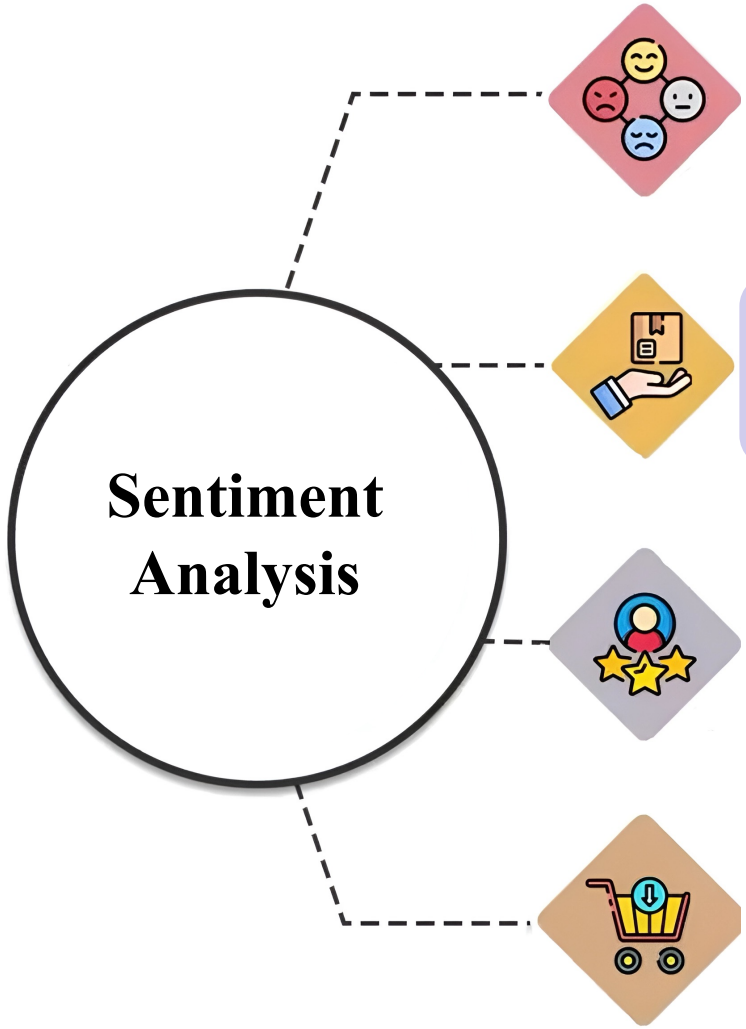
It helps to detect and understand the emotions of the people

I'm so happy to recieve this gift

Happiness

Sentiment Analysis

Types of Sentiment Analysis



Aspect-based Sentiment Analysis
It is more focused on the aspects of a particular prodict or service

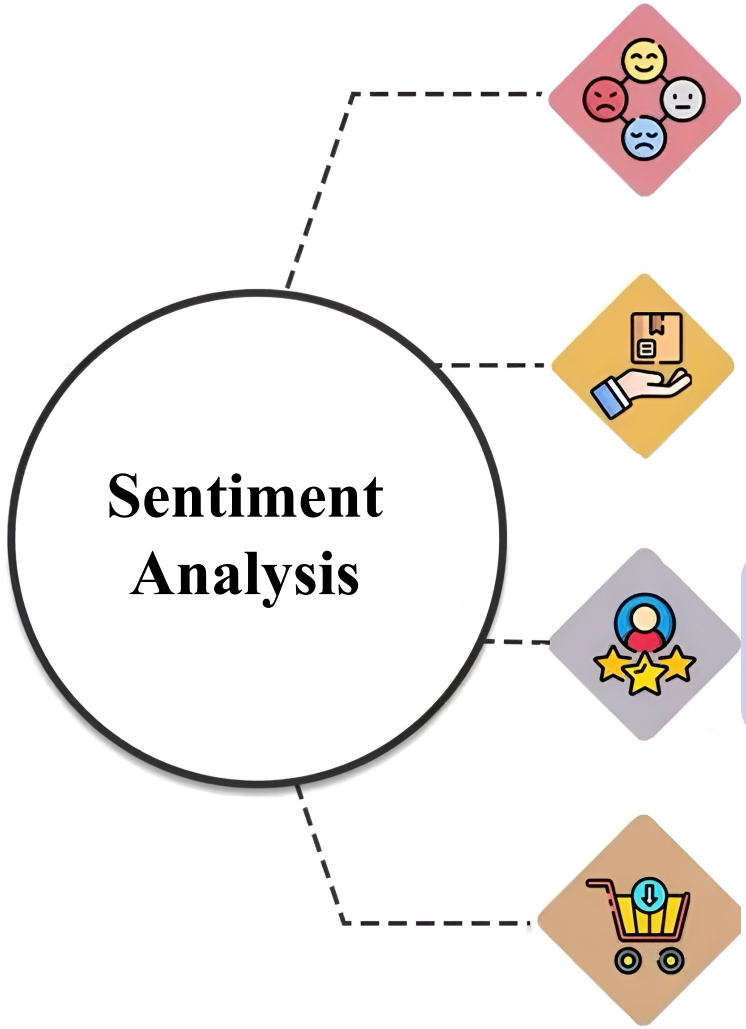
The **food** was delicious, but the **service** was terrible

Positive

Negative

Sentiment Analysis

Types of Sentiment Analysis



The food was okay, but nothing special

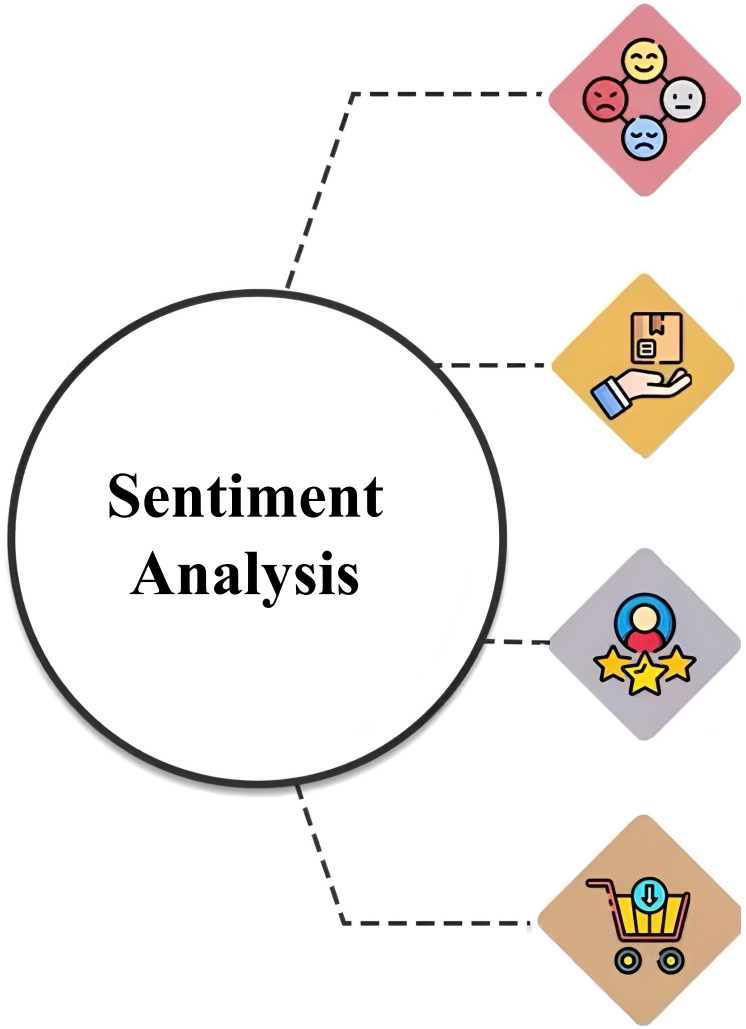
Neural

Fine Grained Sentiment Analysis

It helps in studying the ratings and reviews given by the customers

Sentiment Analysis

Types of Sentiment Analysis



The product arrived damaged. I want a replacement

- Negative
- Complaint

How can I open a savings account?

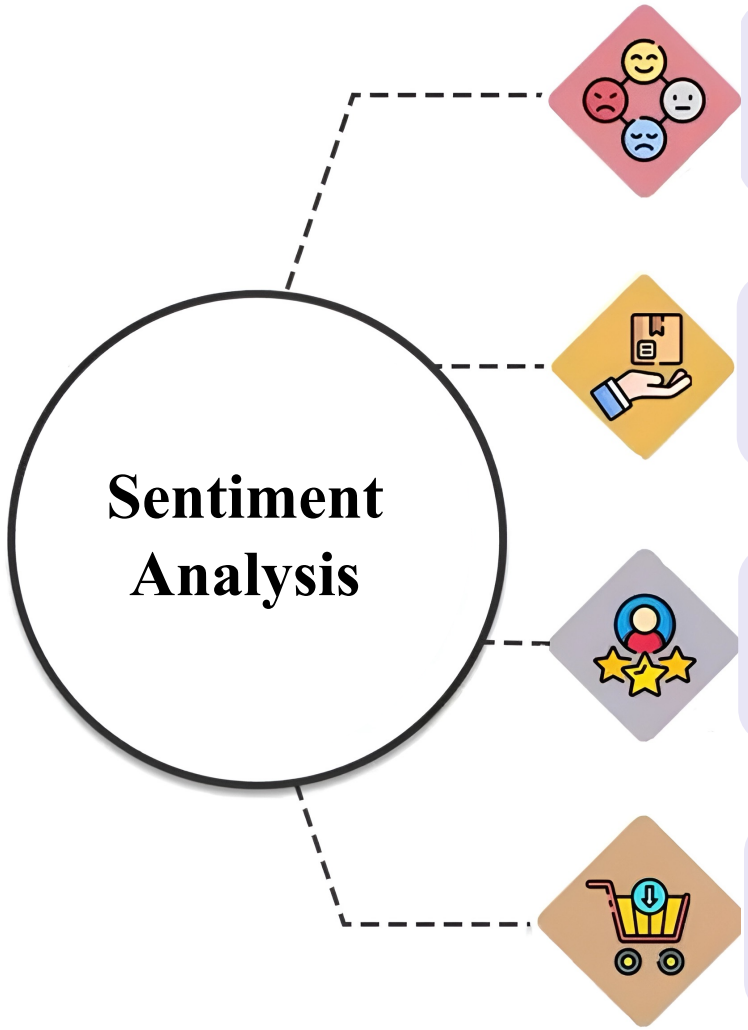
- Neutral
- Information Request

Intent-based Sentiment Analysis

To know the intent of the customers

Sentiment Analysis

Types of Sentiment Analysis



Emotion Detection Sentiment Analysis
It helps to detect and understand the emotions of the people

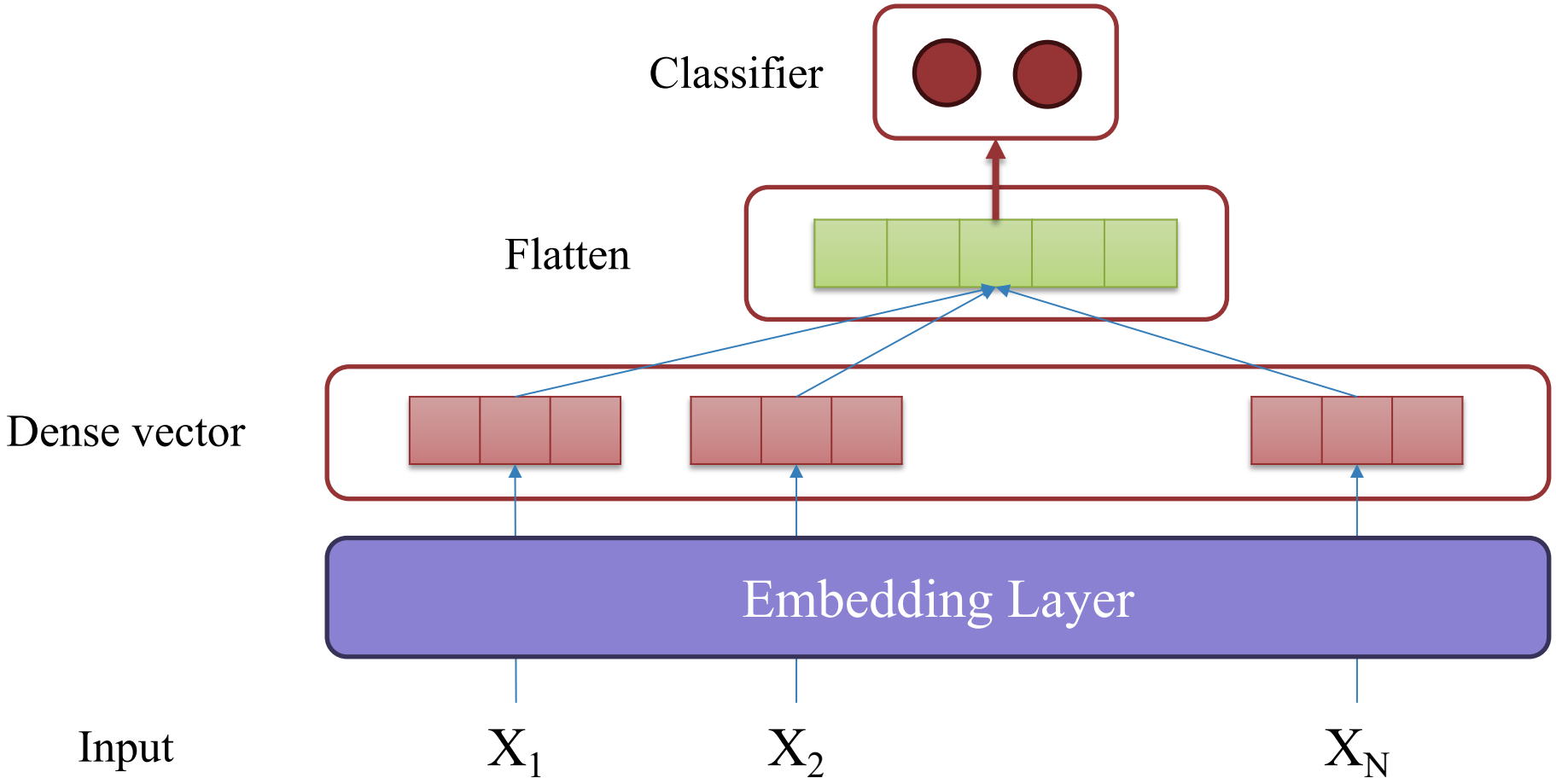
Aspect-based Sentiment Analysis
It is more focused on the aspects of a particular prodict or service

Fine Grained Sentiment Analysis
It helps in studying the ratings and reviews given by the customers

Intent-based Sentiment Analysis
To know the intent of the customers

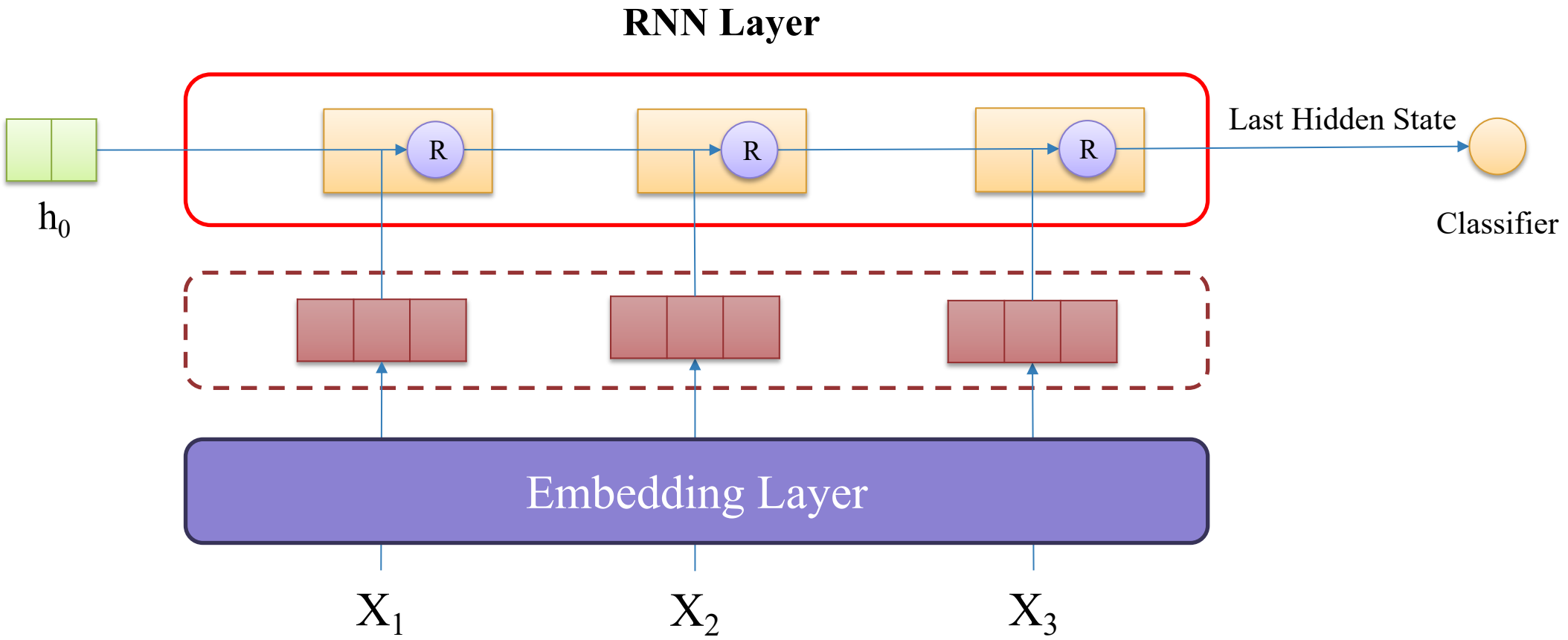
Sentiment Analysis

Sentiment Analysis using Neural Network



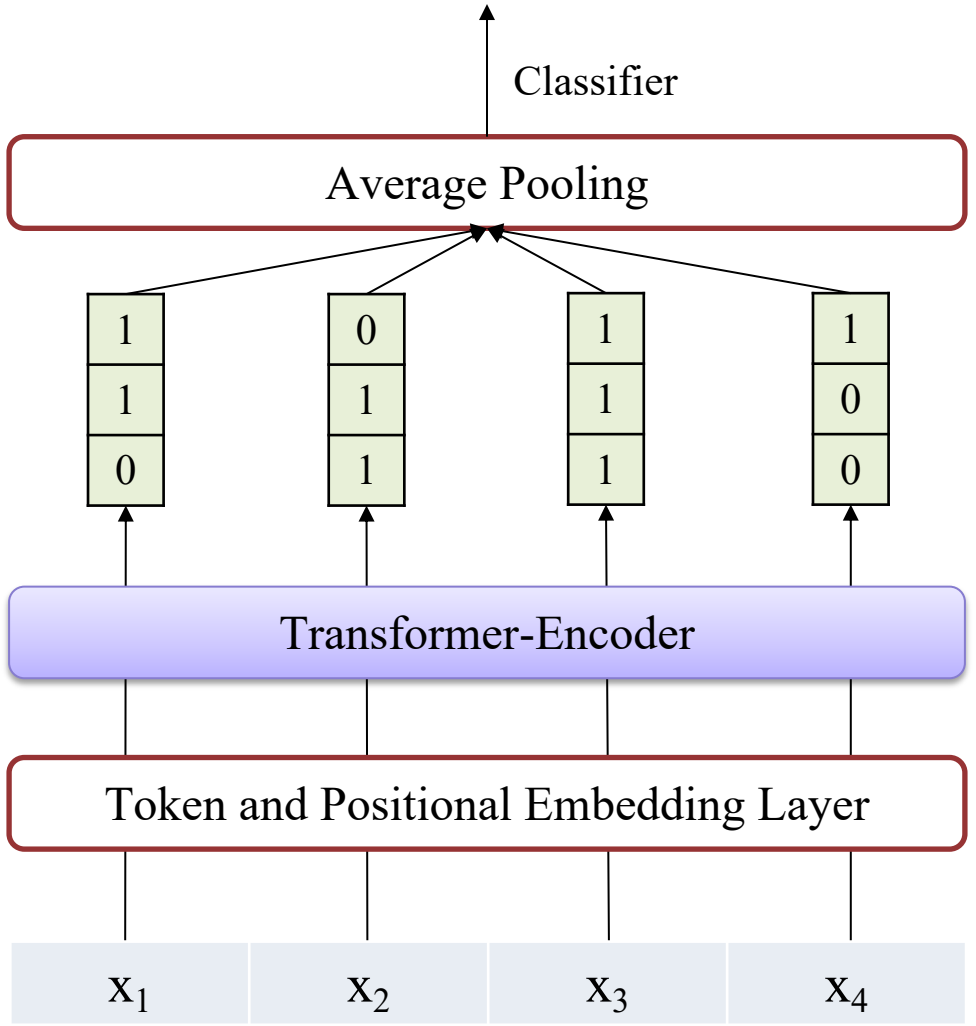
Sentiment Analysis

Sentiment Analysis using RNNs



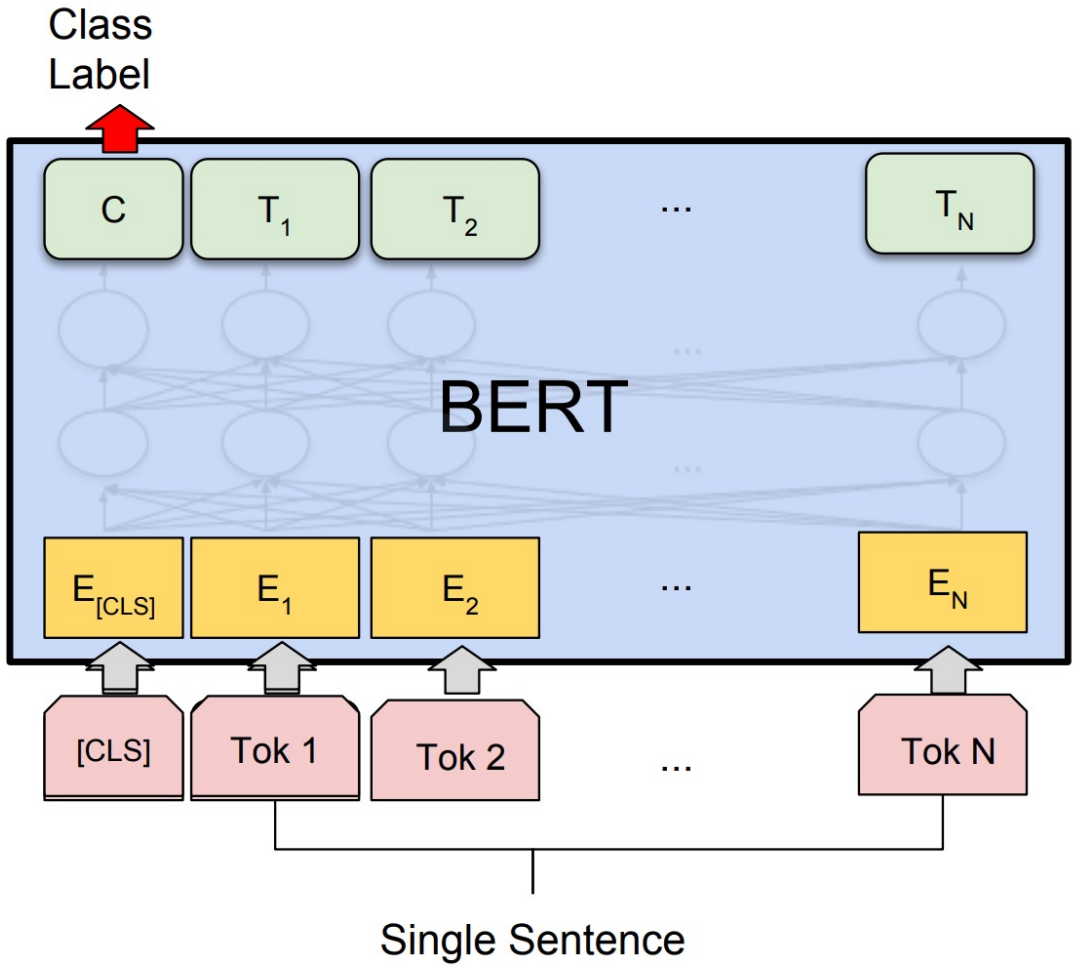
Sentiment Analysis

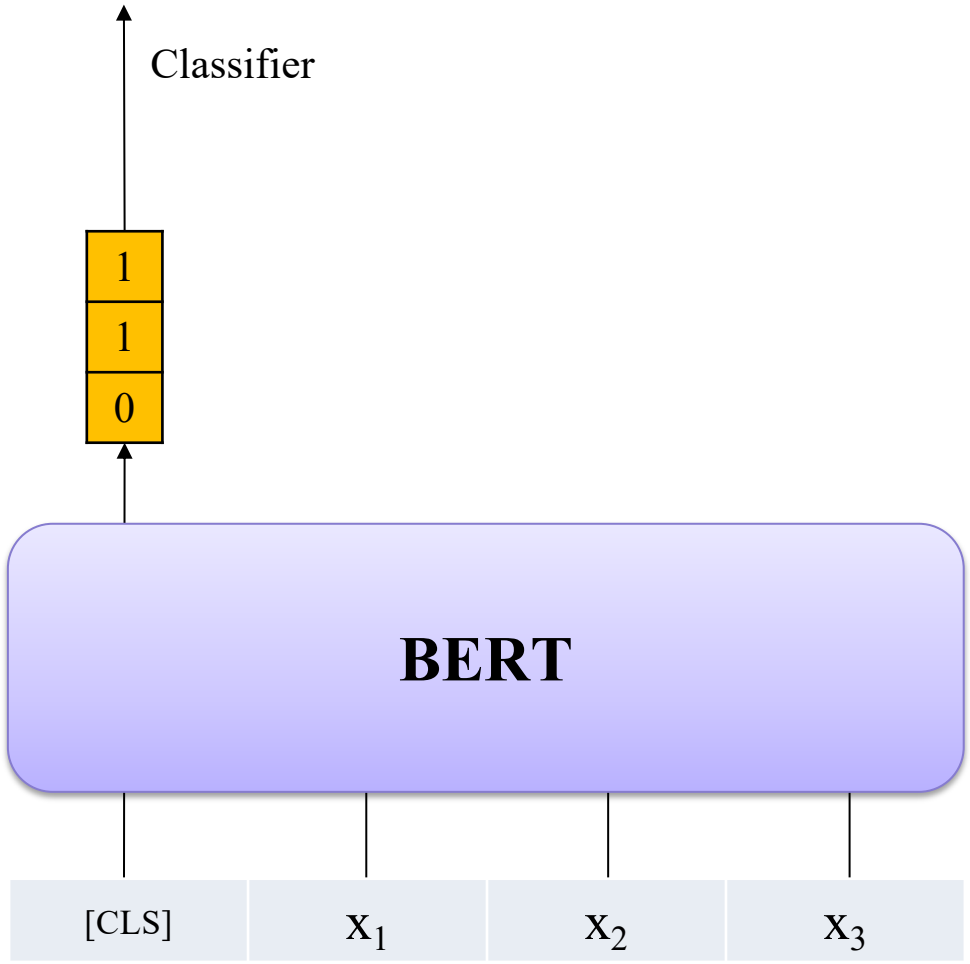
Sentiment Analysis using Transformer-Encoder



Sentiment Analysis

Sentiment Analysis using Pre-trained LMs





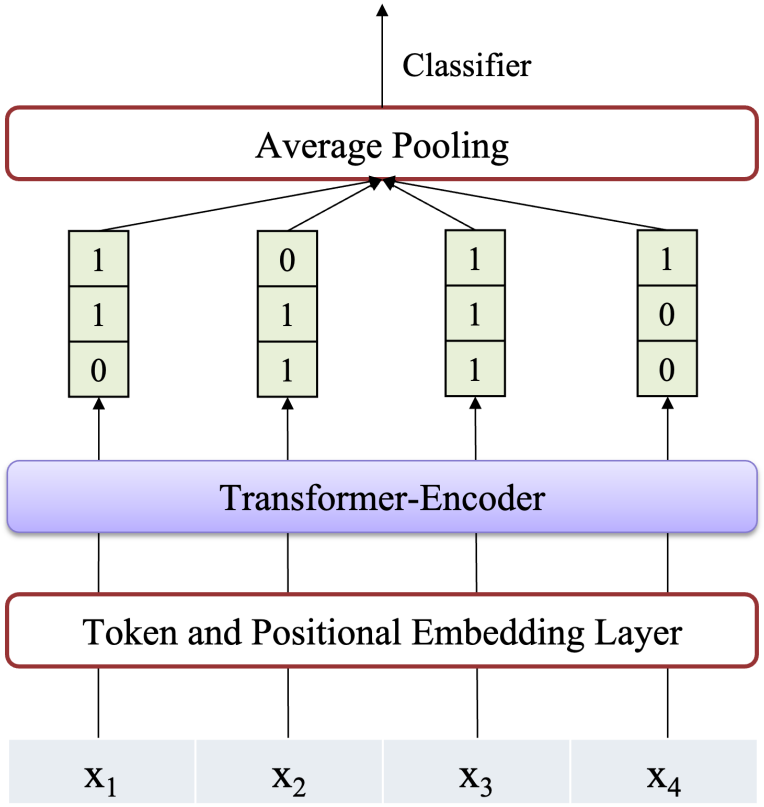
Outline

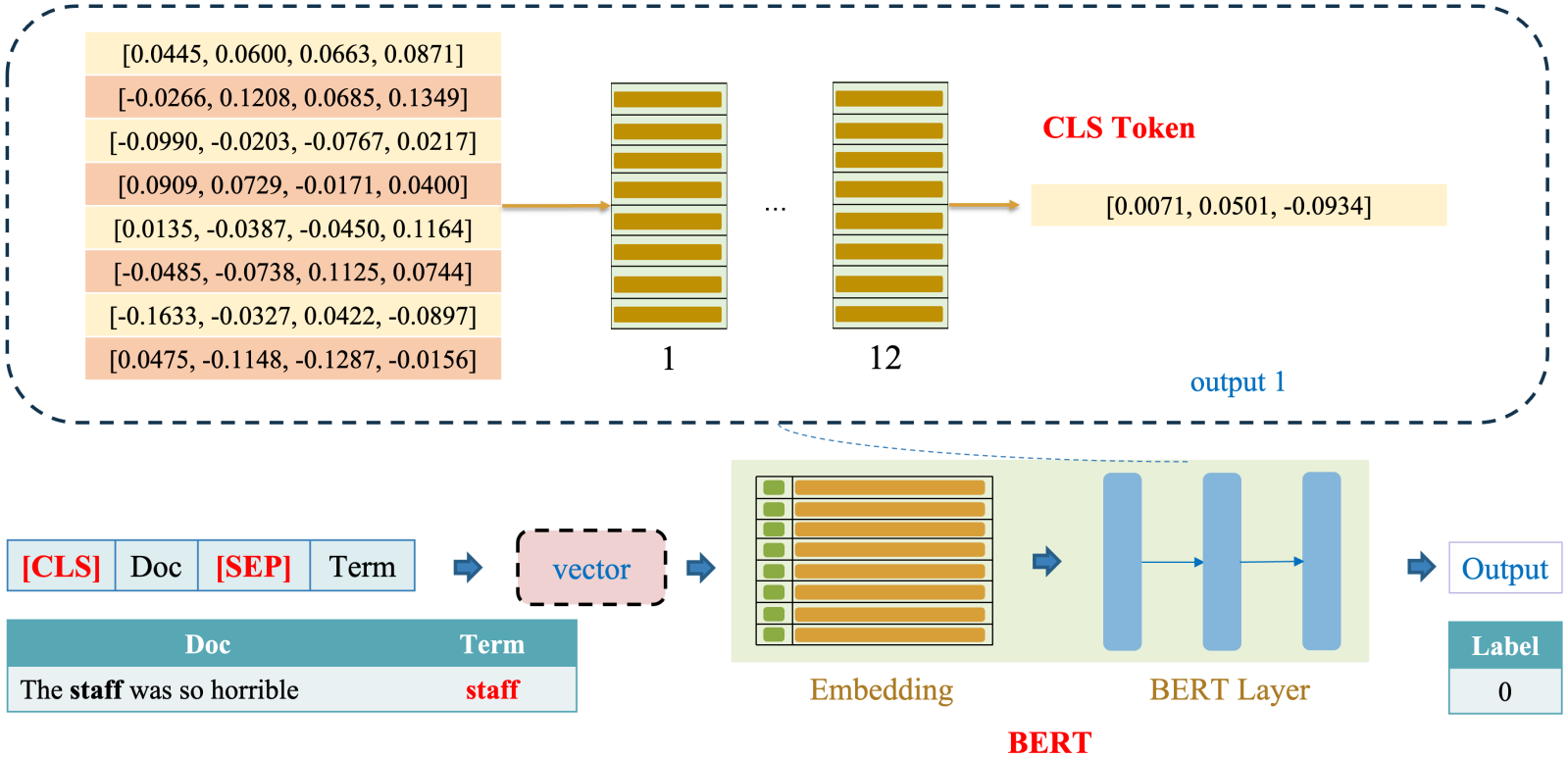
SECTION 1

Sentiment Analysis

SECTION 2

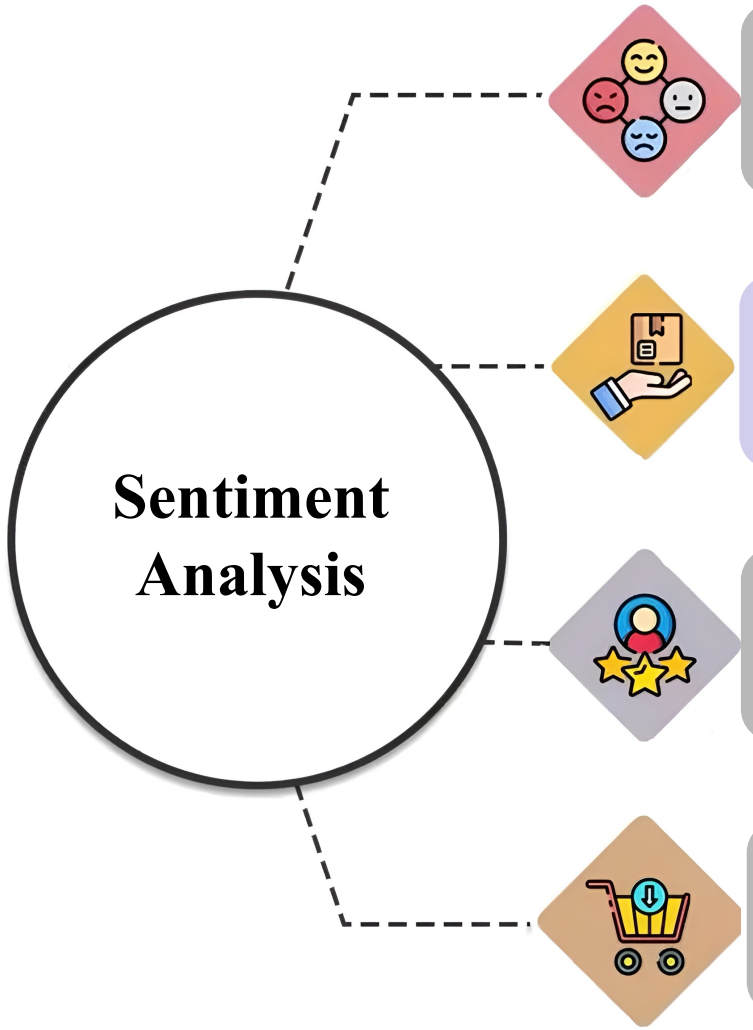
Aspect-based SA





Sentiment Analysis

Types of Sentiment Analysis



Emotion Detection Sentiment Analysis
It helps to detect and understand the emotions of the people

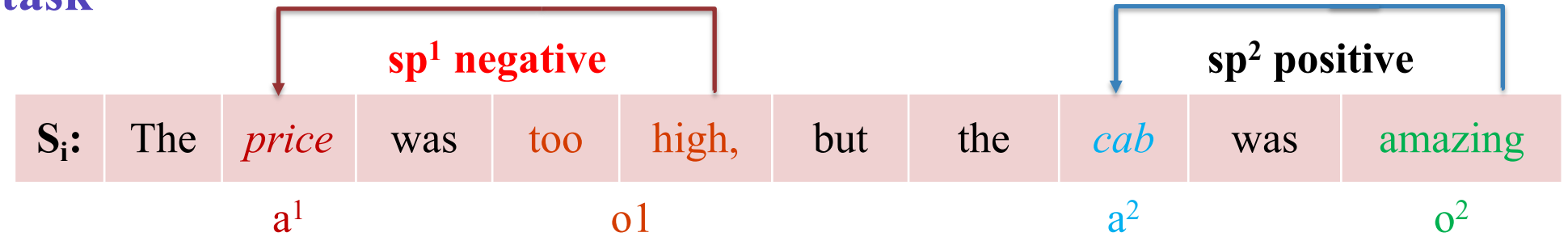
Aspect-based Sentiment Analysis
It is more focused on the aspects of a particular prodict or service

Fine Grained Sentiment Analysis
It helps in studying the ratings and reviews given by the customers

Intent-based Sentiment Analysis
To know the intent of the customers

Aspect-based Sentiment Analysis

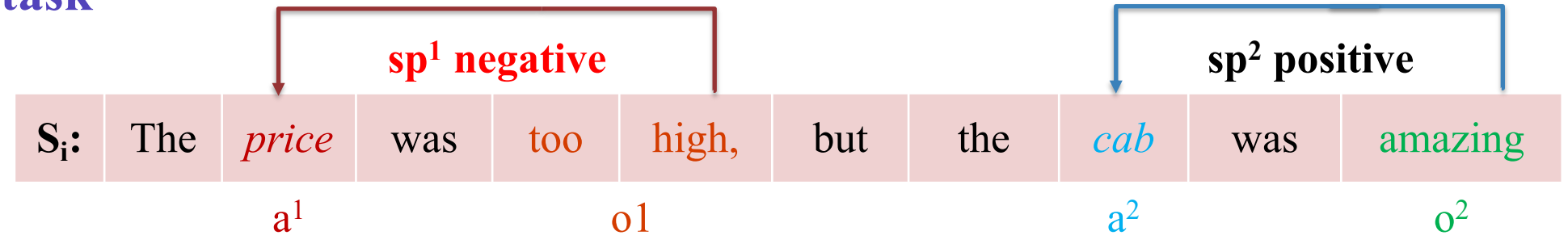
Subtask



- Aspect Term Extraction (ATE)
- Aspect Term Sentiment Classification (ATSC)
- Aspect Sentiment Pair Extraction (ASPE)
- Aspect Oriented Opinion Extraction (AOOE)
- Aspect Opinion Pair Extraction (AOPE)
- Aspect Opinion Sentiment Triplet Extraction (AOSTE)

Aspect-based Sentiment Analysis

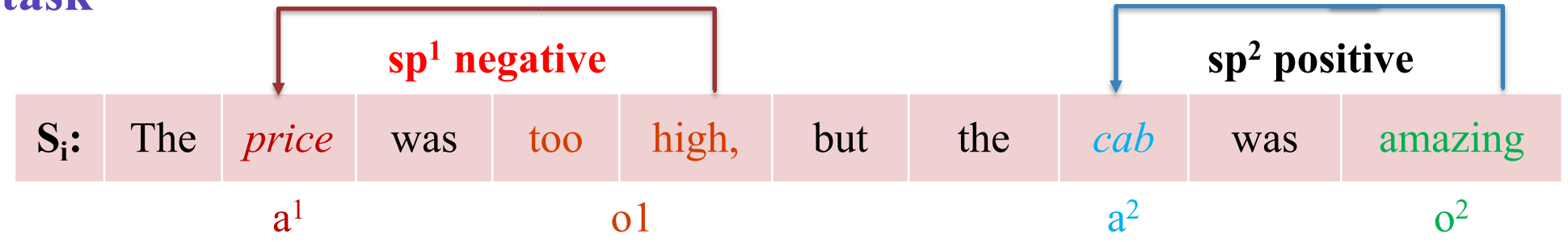
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)		
Aspect Sentiment Pair Extraction (ASPE)		
Aspect Oriented Opinion Extraction (AOOE)		
Aspect Opinion Pair Extraction (AOPE)		
Aspect Opinion Sentiment Triplet Extraction (AOSTE)		

Aspect-based Sentiment Analysis

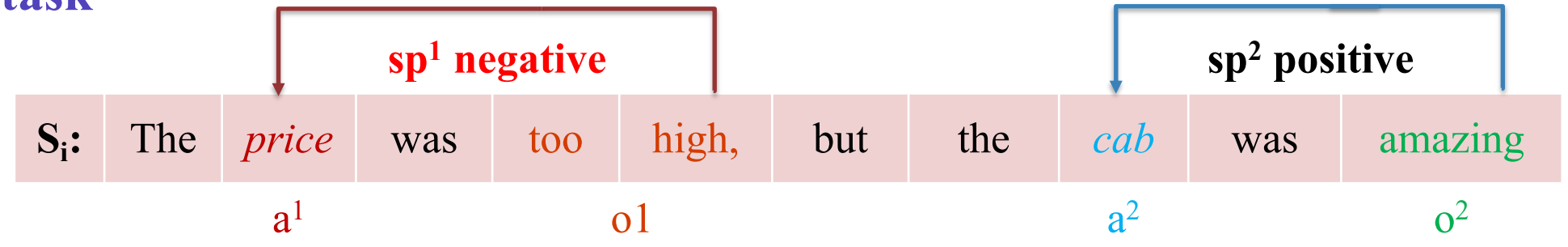
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1$, $S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)		
Aspect Oriented Opinion Extraction (AOOE)		
Aspect Opinion Pair Extraction (AOPE)		
Aspect Opinion Sentiment Triplet Extraction (AOSTE)		

Aspect-based Sentiment Analysis

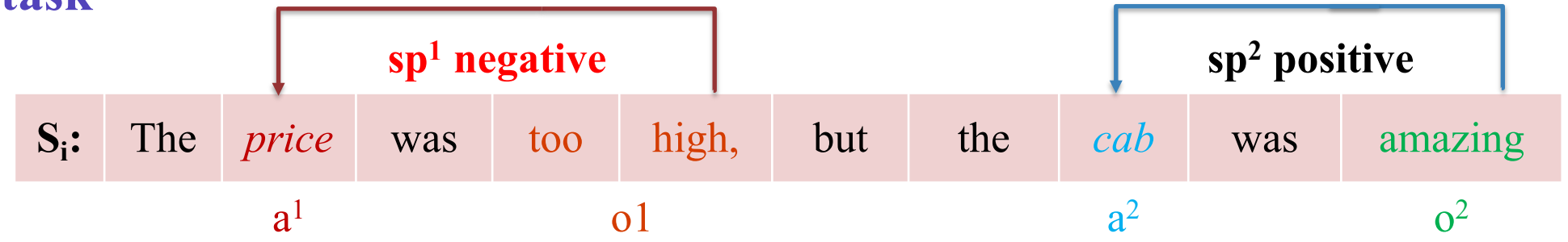
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1$, $S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)	S_i	$(a^1, sp^1), (a^2, sp^2)$
Aspect Oriented Opinion Extraction (AOOE)		
Aspect Opinion Pair Extraction (AOPE)		
Aspect Opinion Sentiment Triplet Extraction (AOSTE)		

Aspect-based Sentiment Analysis

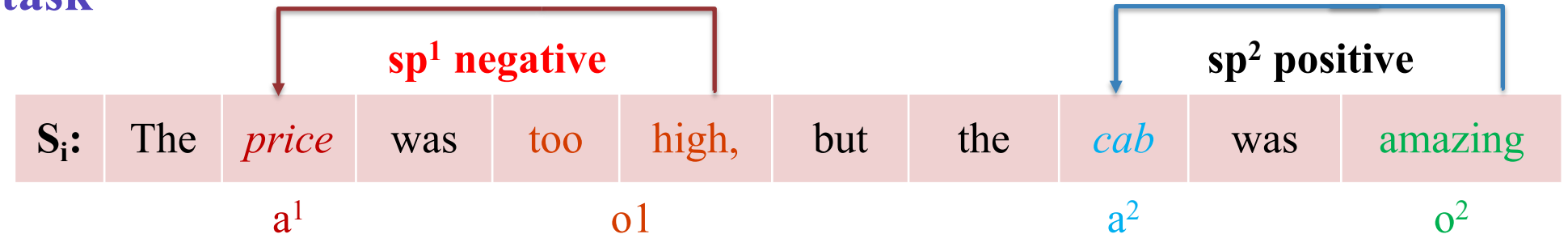
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1$, $S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)	S_i	$(a^1, sp^1), (a^2, sp^2)$
Aspect Oriented Opinion Extraction (AOOE)	$S_i + a^1$, $S_i + a^2$	o^1, o^2
Aspect Opinion Pair Extraction (AOPE)		
Aspect Opinion Sentiment Triplet Extraction (AOSTE)		

Aspect-based Sentiment Analysis

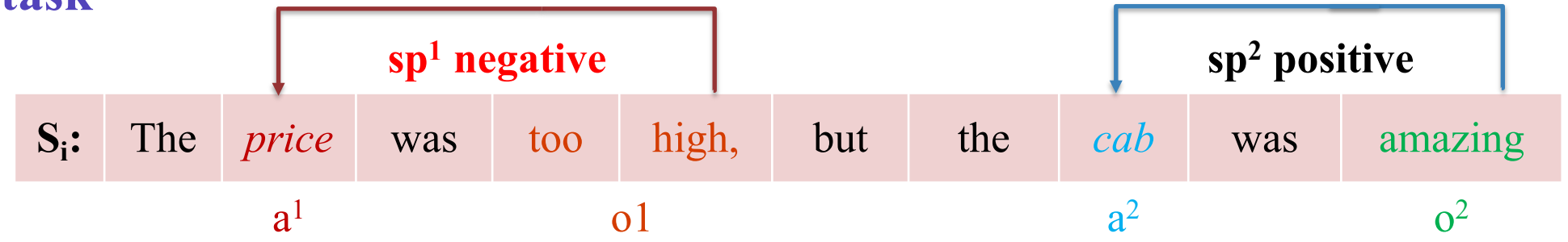
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1, S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)	S_i	$(a^1, sp^1), (a^2, sp^2)$
Aspect Oriented Opinion Extraction (AOOE)	$S_i + a^1, S_i + a^2$	o^1, o^2
Aspect Opinion Pair Extraction (AOPE)	S_i	$(a^1, o^1), (a^2, o2)$
Aspect Opinion Sentiment Triplet Extraction (AOSTE)		

Aspect-based Sentiment Analysis

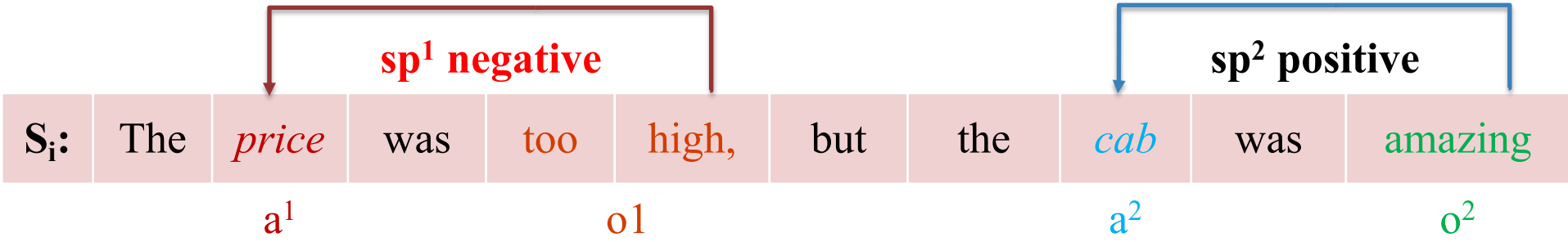
Subtask



Aspect Term Extraction (ATE)	S_i	a^1, a^2
Aspect Term Sentiment Classification (ATSC)	$S_i + a^1, S_i + a^2$	sp^1, sp^2
Aspect Sentiment Pair Extraction (ASPE)	S_i	$(a^1, sp^1), (a^2, sp^2)$
Aspect Oriented Opinion Extraction (AOOE)	$S_i + a^1, S_i + a^2$	o^1, o^2
Aspect Opinion Pair Extraction (AOPE)	S_i	$(a^1, o^1), (a^2, o2)$
Aspect Opinion Sentiment Triplet Extraction (AOSTE)	S_i	$(a^1, o^1, sp^1), (a^2, o^2, sp^2)$

Aspect Sentiment Pair Extraction

Example



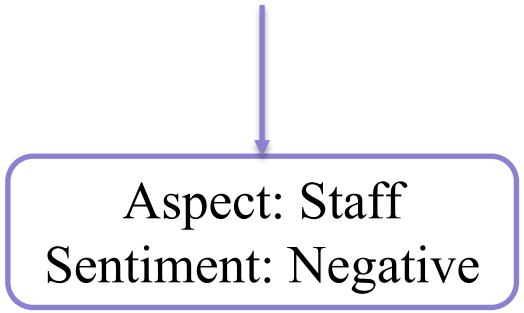
S_i	$(a^1, sp^1), (a^2, sp^2)$

Aspect Sentiment Pair Extraction

SemEval Task 4 Dataset

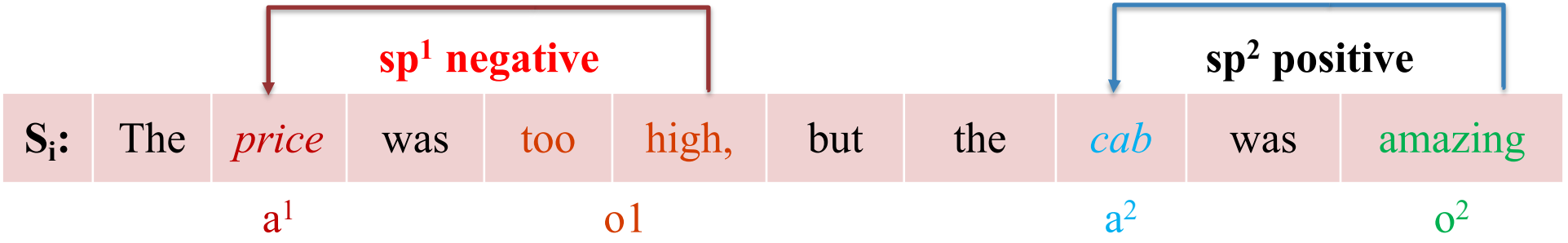
Domain	Train	Test	Total
restaurants	3041	800	3841
laptops	3045	800	3845
Total	6086	1600	7686

But	the	staff	was	so	horribale	to	us

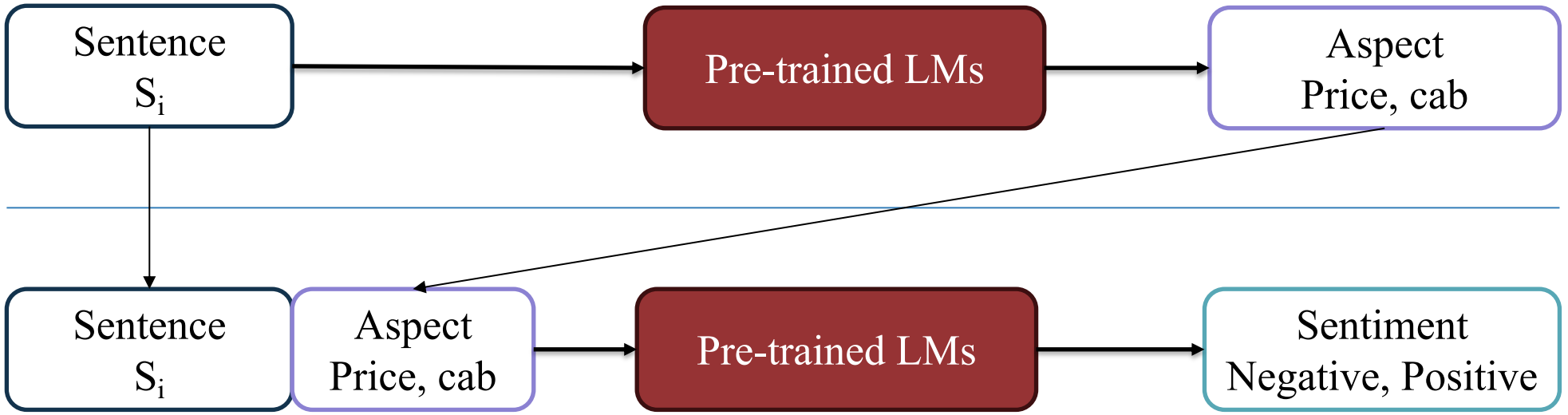


Aspect Sentiment Pair Extraction

Approach



Stage 1: Aspect Term Extraction (Token-level Classification)



Stage 2: Aspect Term Sentiment Extraction (Document-level Classification)

Aspect Sentiment Pair Extraction

Approach

But	the	staff	was	so	horribale	to	us

Stage 1: Aspect Term Extraction (Token-level Classification)

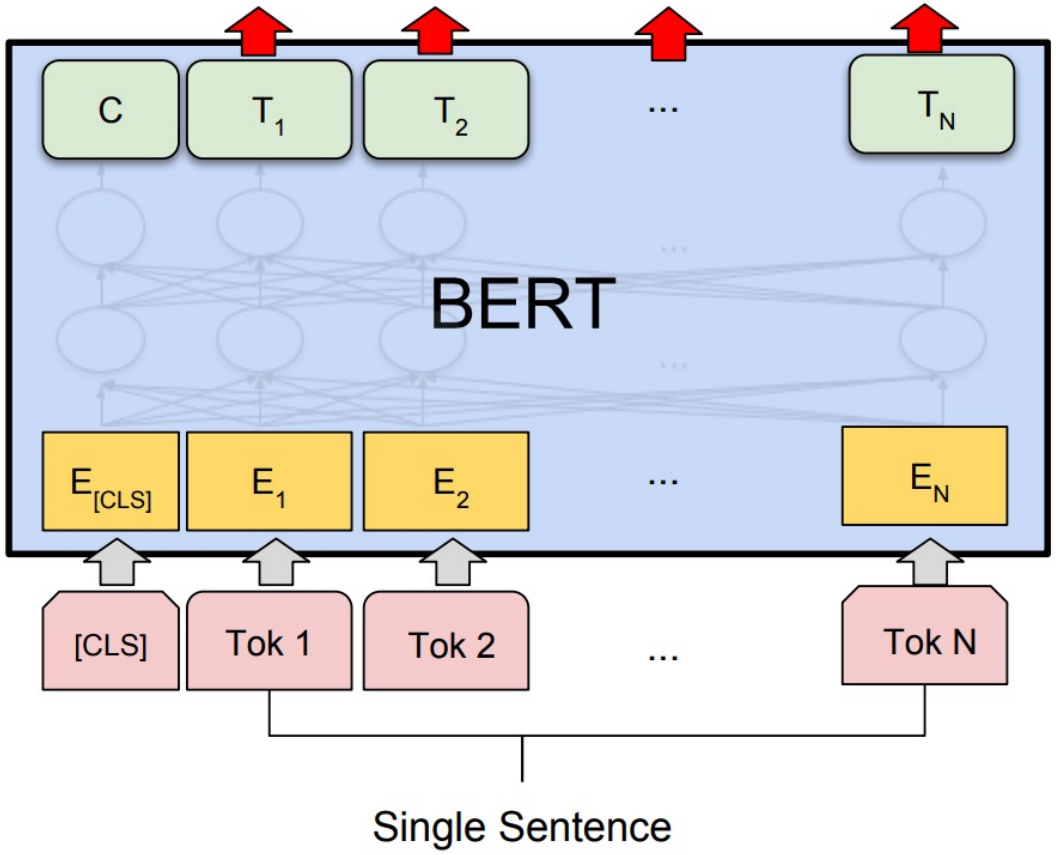
0	0	1	0	0	0	0	0

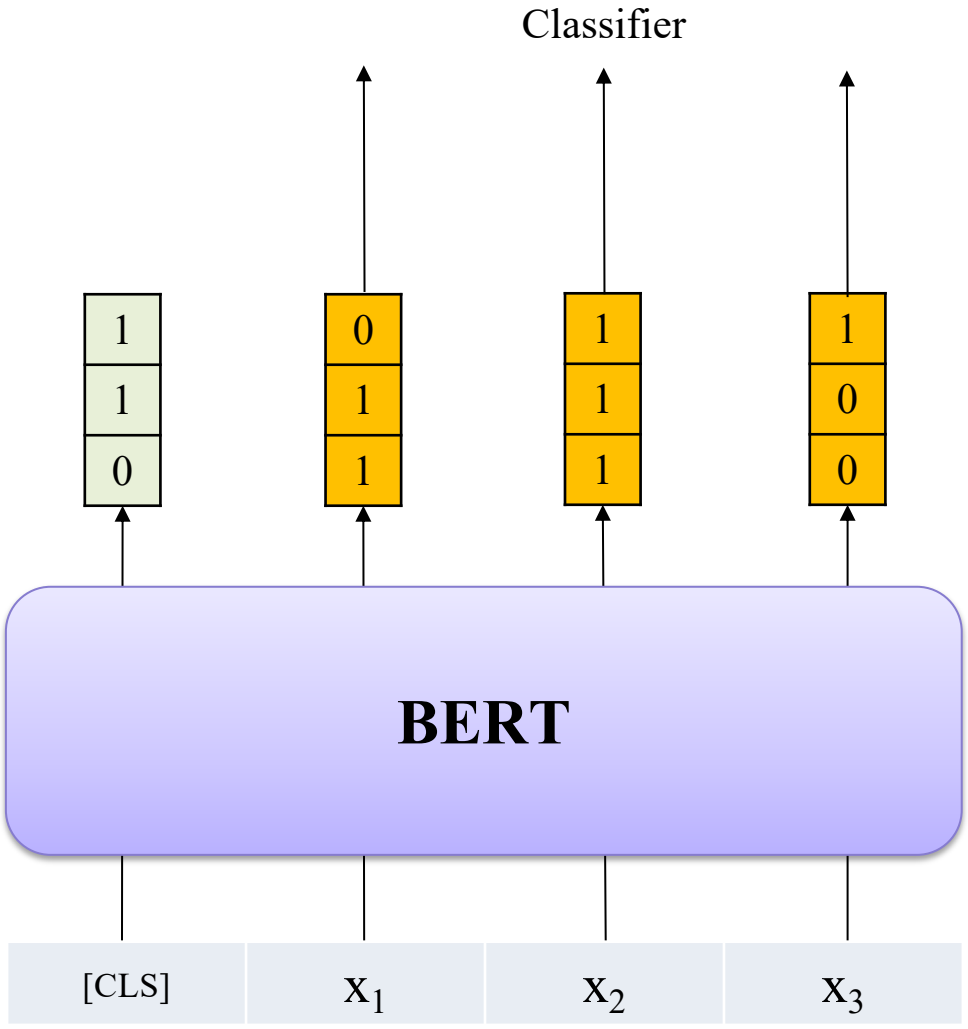
Stage 2: Aspect Term Sentiment Extraction (Document-level Classification)

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

Aspect Sentiment Pair Extraction

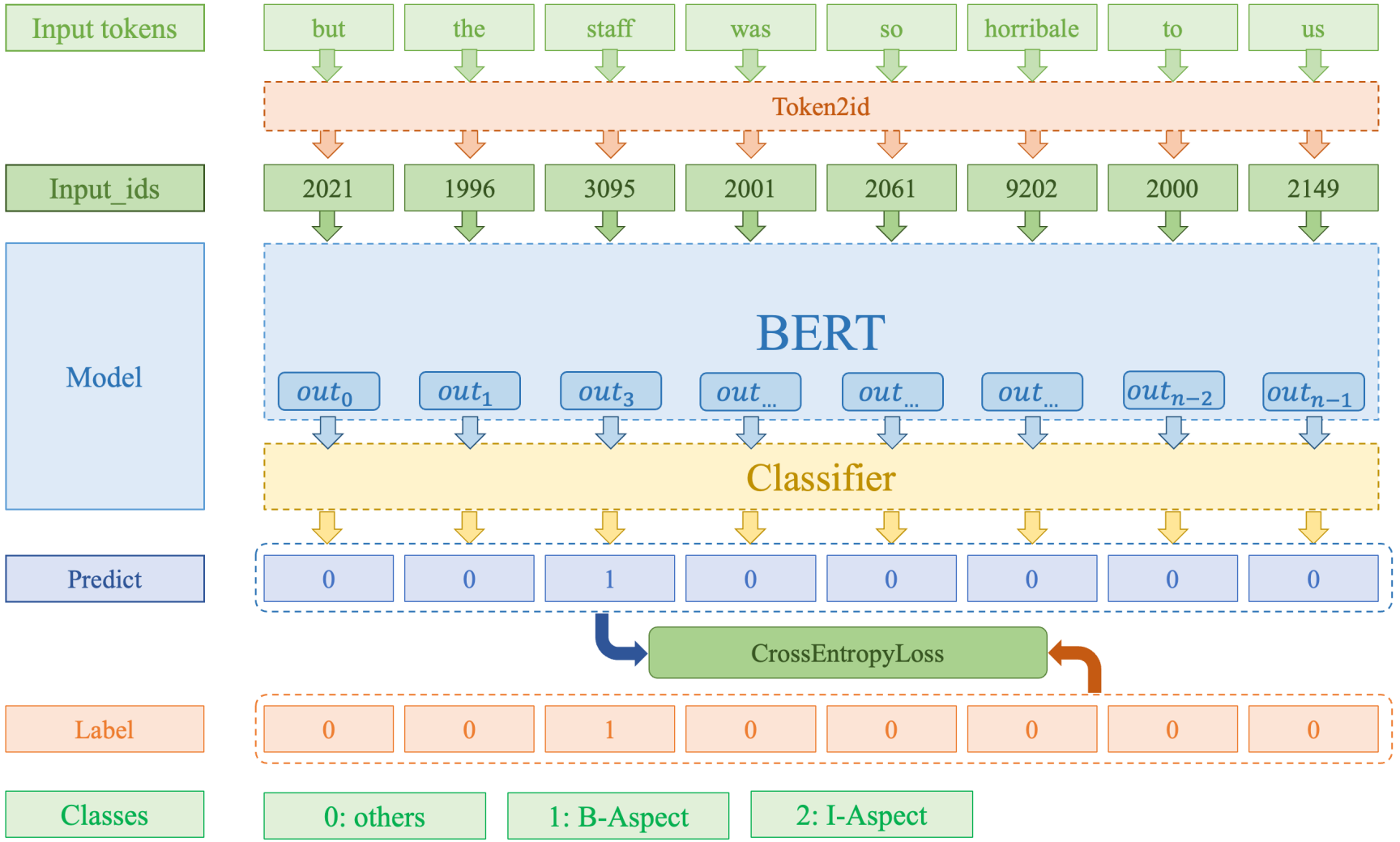
Aspect Term Extraction





Aspect Sentiment Pair Extraction

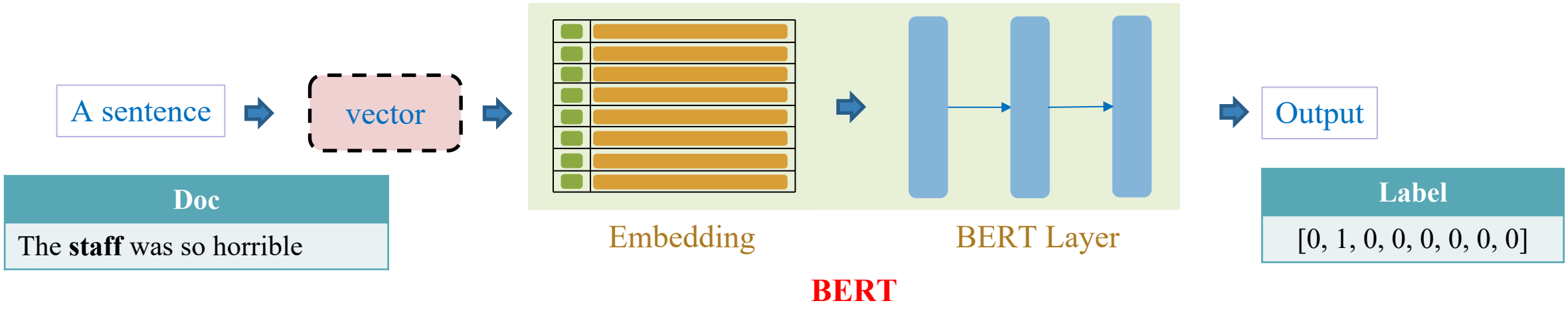
Aspect Term Extraction



Aspect Term Extraction

Doc	Label
The **staff** was so horrible	[0, **1**, 0, 0, 0, 0, 0, 0]
The **bread** is top notch as well	[0, **1**, 0, 0, 0, 0, 0, 0]

TERM



Aspect Term Extraction

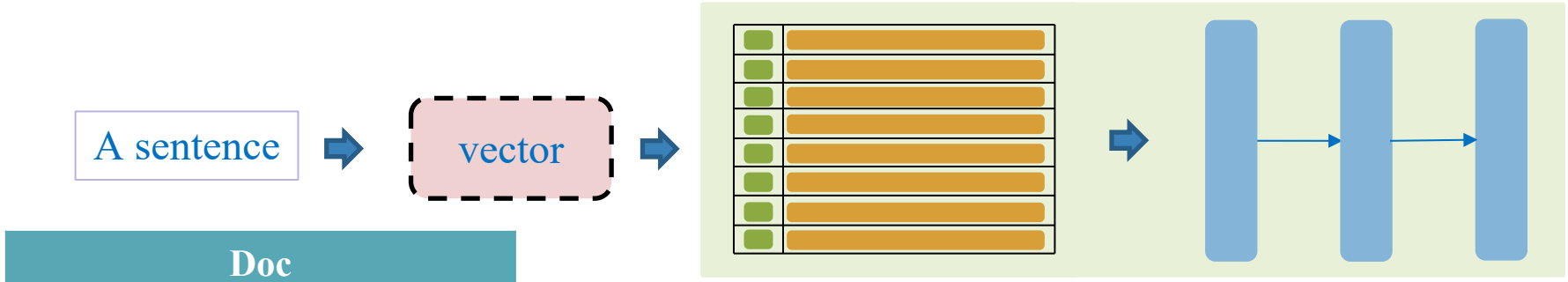
Baseline Models

RNNs Conv1D

Transformer-Encoder

Pretrained Models

BERTs



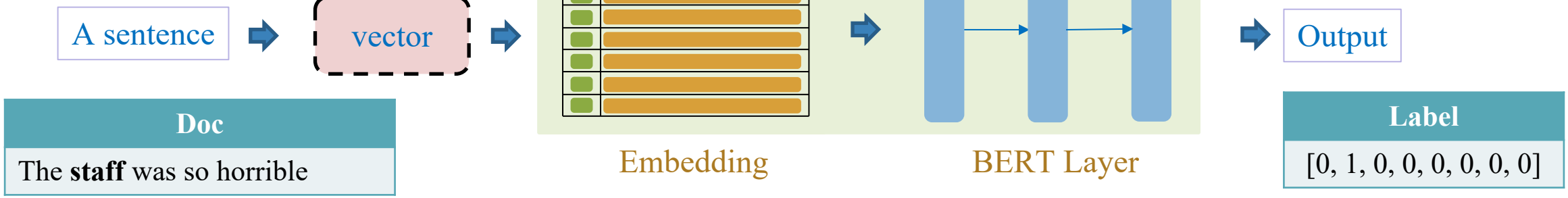
Doc
The **staff** was so horrible

Label
[0, 1, 0, 0, 0, 0, 0, 0]

Aspect Term Extraction

index	token
0	[UNK]
1	[pad]
2	the
3	staff
4	was
5	so
6	horrible
…	…
30522	…

the	2	0
staff	3	1
was	4	0
so	5	0
horrible	6	0
sample 1	vec 1	label 1



Doc
The **staff** was so horrible

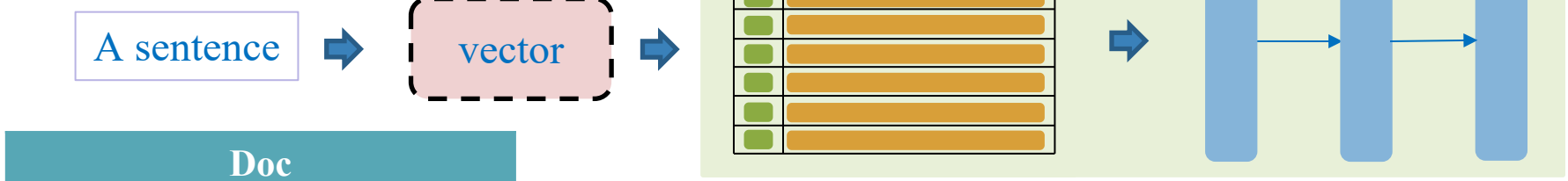
Label
[0, 1, 0, 0, 0, 0, 0, 0]

Aspect Term Extraction

index	token
0	[UNK]
1	[pad]
2	the
3	staff
4	was
5	so
6	horrible
…	…
30522	…

the	2	0
staff	3	1
was	4	0
so	5	0
horrible	6	0
sample 1	vec 1	label 1

```
bert_tokens = []
bert_tags = []
for i in range(len(tokens)):
    t = tokenizer.tokenize(tokens[i])
    bert_tokens += t
    bert_tags += [int(tags[i])]*len(t)

bert_ids = tokenizer.convert_tokens_to_ids(bert_tokens)
```



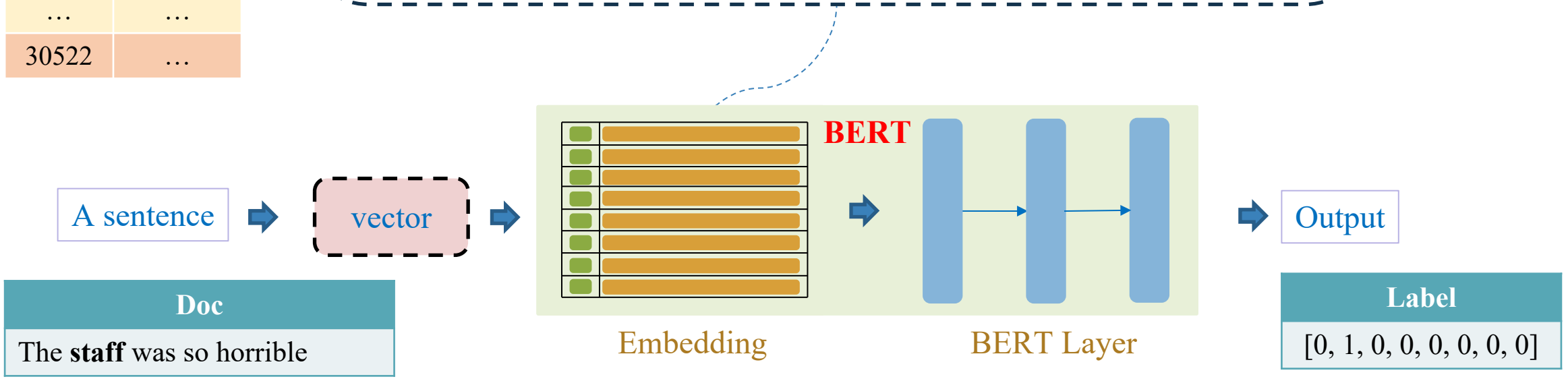
Doc
The **staff** was so horrible

Label
[0, 1, 0, 0, 0, 0, 0, 0]

Aspect Term Extraction

index	token
0	[UNK]
1	[pad]
2	the
3	staff
4	was
5	so
6	horrible
…	…
30522	…

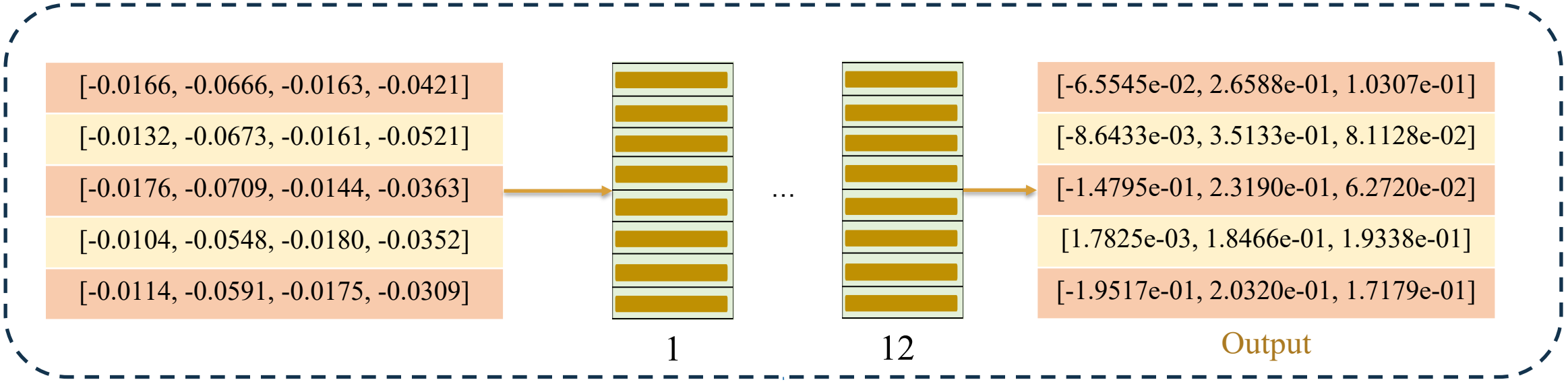
the	2	0	[-0.0166, -0.0666, -0.0163, -0.0421]
staff	3	1	[-0.0132, -0.0673, -0.0161, -0.0521]
was	4	0	[-0.0176, -0.0709, -0.0144, -0.0363]
so	5	0	[-0.0104, -0.0548, -0.0180, -0.0352]
horrible	6	0	[-0.0114, -0.0591, -0.0175, -0.0309]
sample 1	vec 1	label 1	

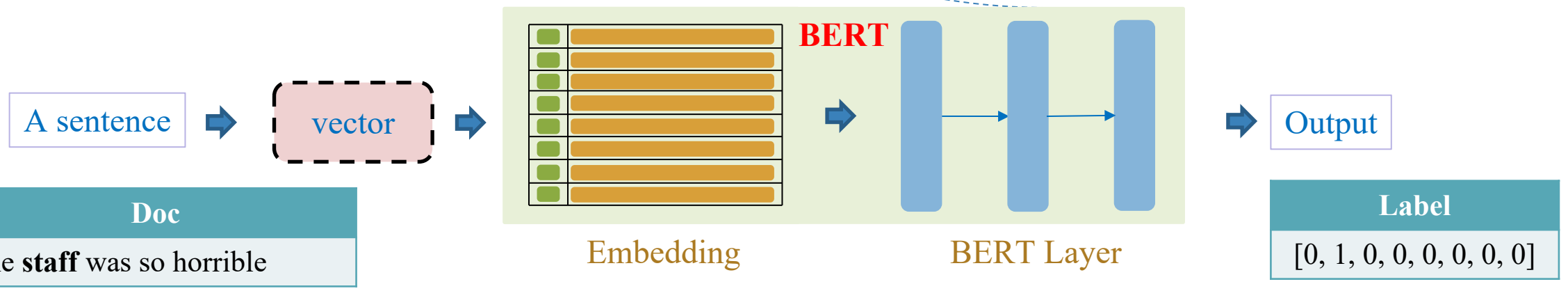


Doc
The **staff** was so horrible

Label
[0, 1, 0, 0, 0, 0, 0, 0]

Aspect Term Extraction





Doc
The **staff** was so horrible

Label
[0, 1, 0, 0, 0, 0, 0, 0]

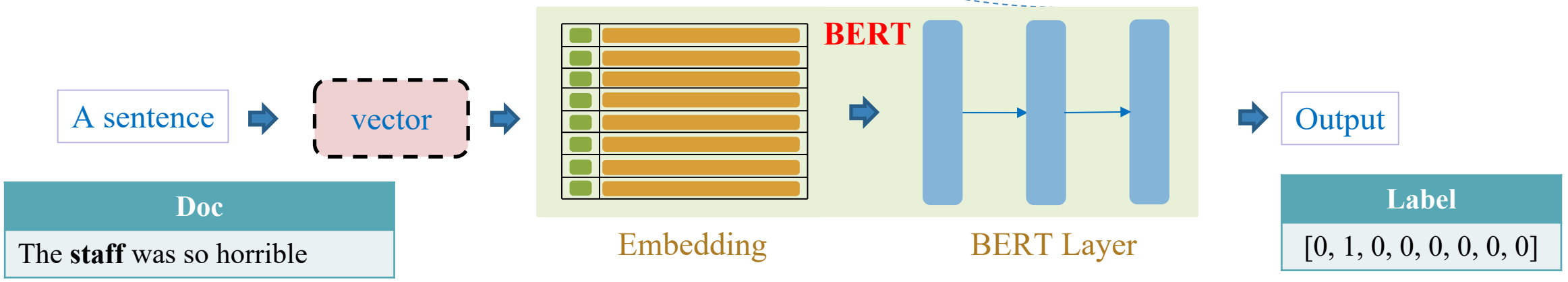
Aspect Term Extraction

```python
from transformers import AutoModelForTokenClassification

model = AutoModelForTokenClassification.from_pretrained(
    "distilbert/distilbert-base-uncased",
    num_labels=3, id2label=id2label, label2id=label2id
)
```

```python
preds = model(
    torch.tensor([preprocessed_ds["train"][0]['input_ids']]),
    labels=torch.tensor([preprocessed_ds["train"][0]['labels']])
)
```

[-6.5545e-02, 2.6588e-01, 1.0307e-01]
[-8.6433e-03, 3.5133e-01, 8.1128e-02]
[-1.4795e-01, 2.3190e-01, 6.2720e-02]
[1.7825e-03, 1.8466e-01, 1.9338e-01]
[-1.9517e-01, 2.0320e-01, 1.7179e-01]

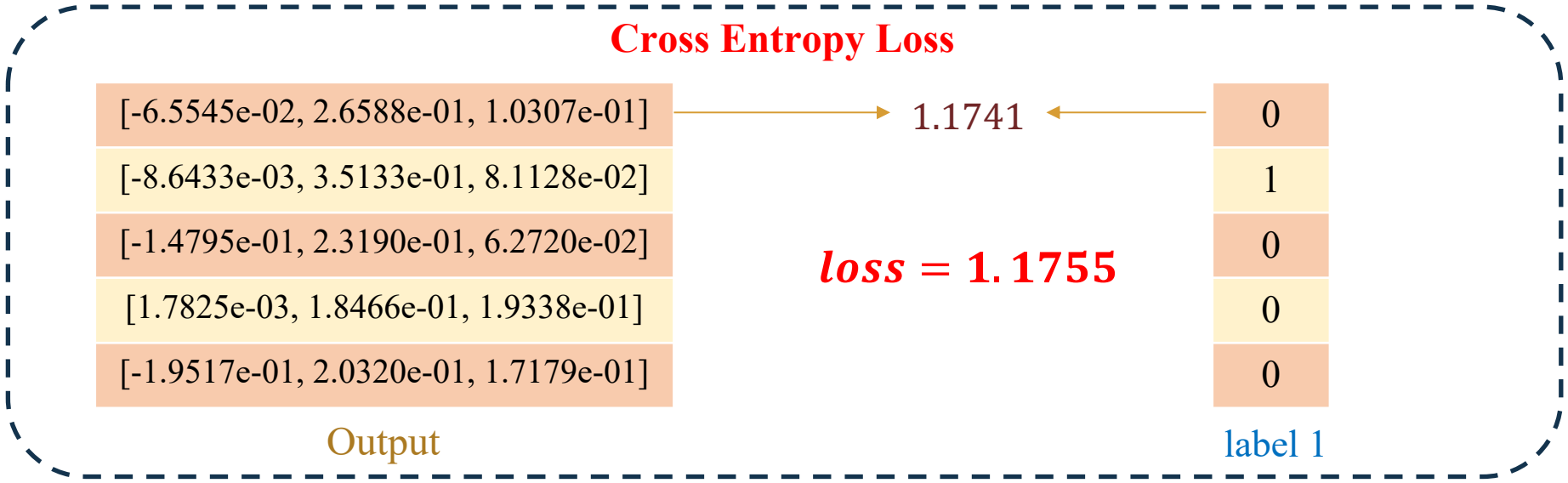
Output

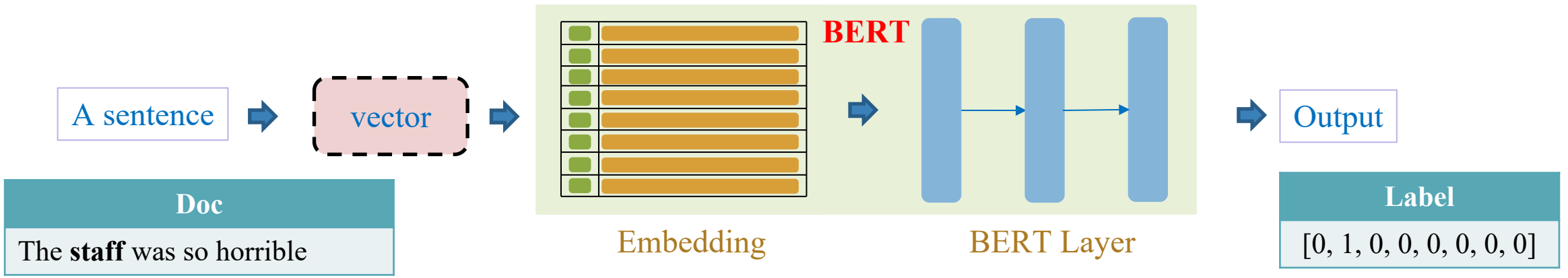


Doc
The **staff** was so horrible

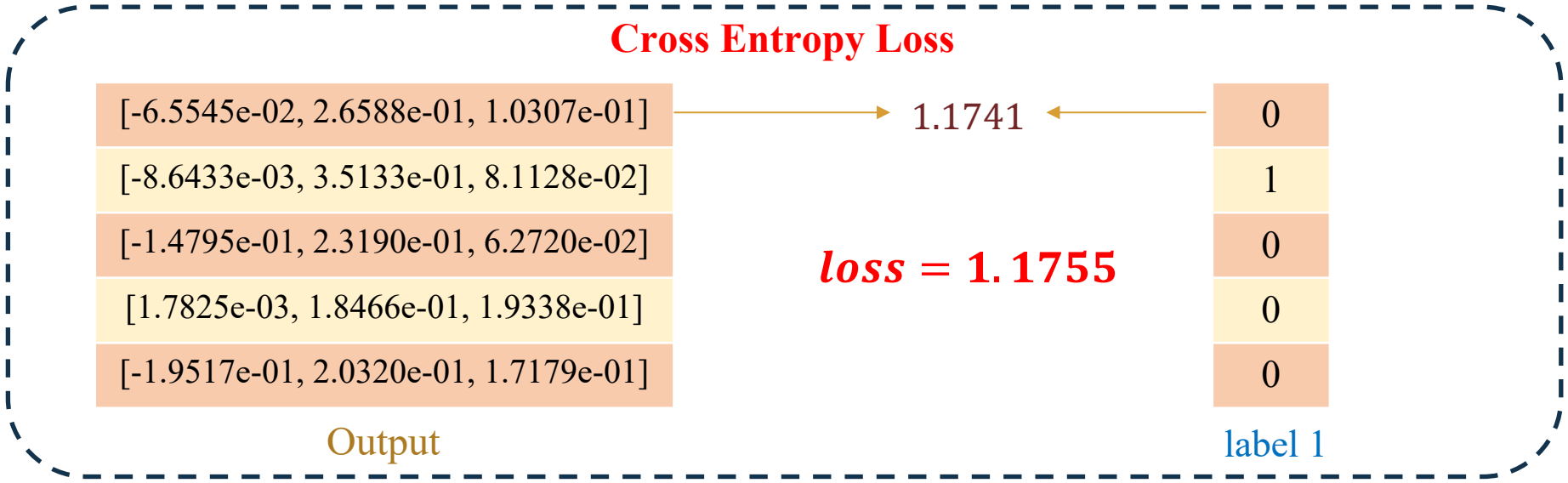
Label
[0, 1, 0, 0, 0, 0, 0, 0]

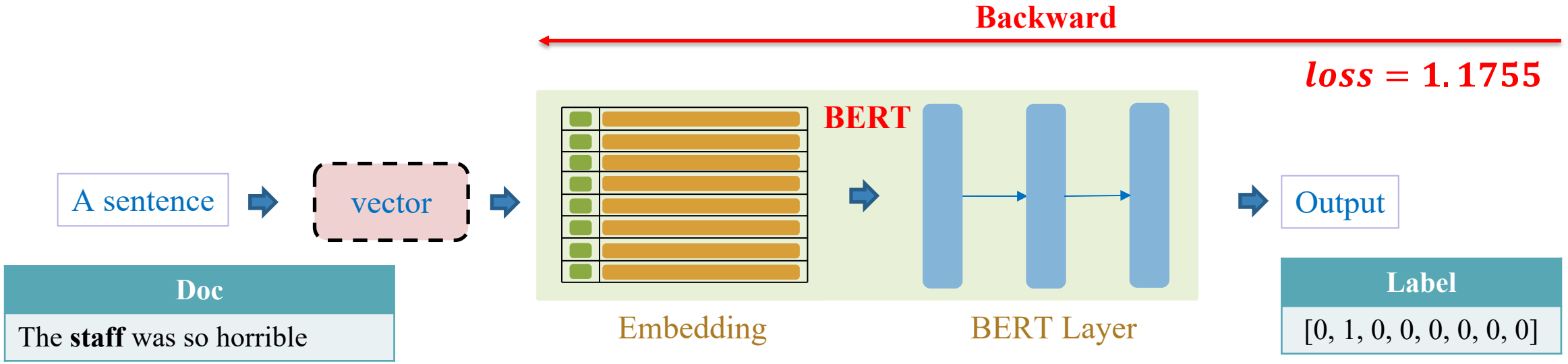
Aspect Term Extraction





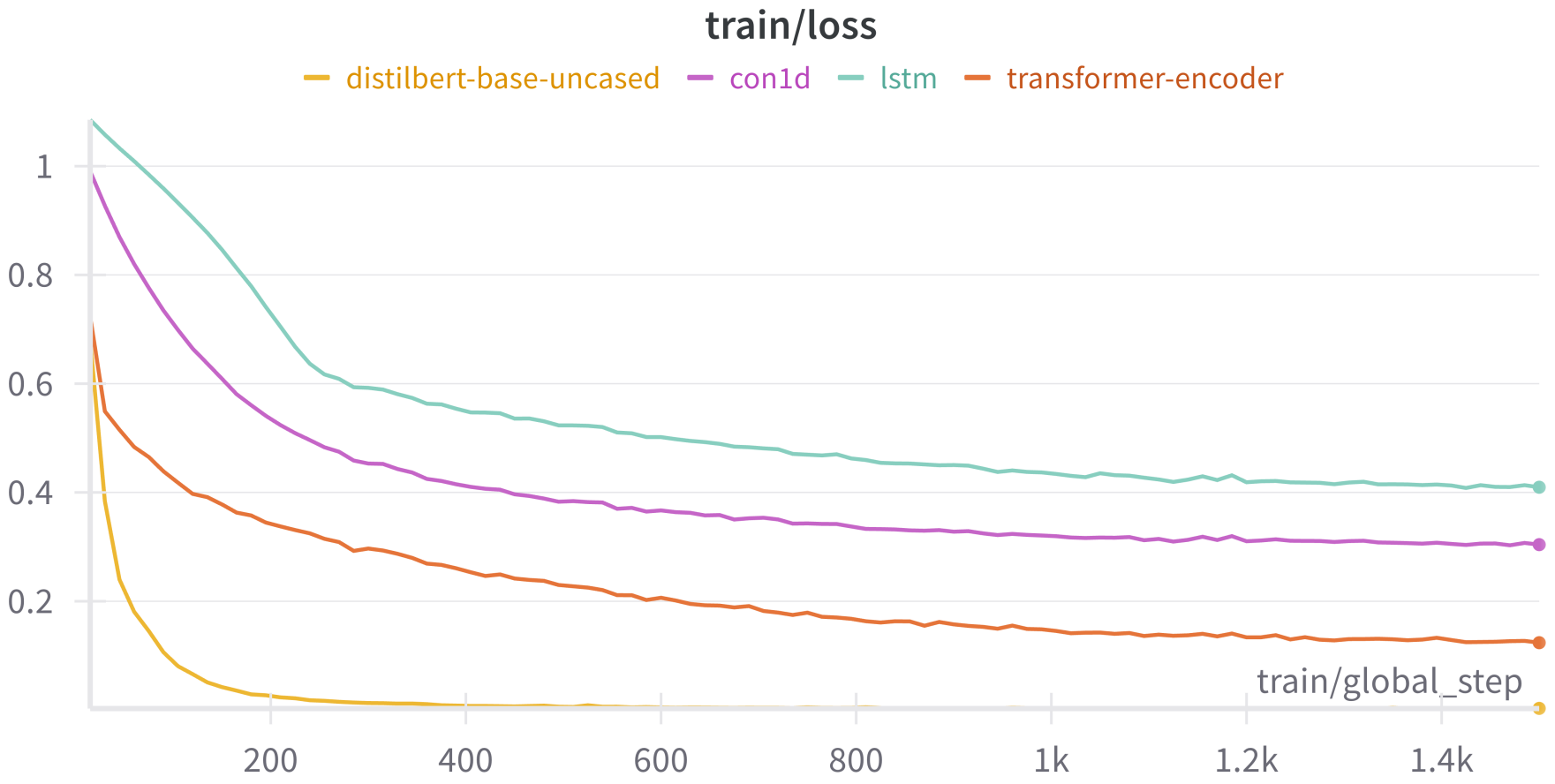
Aspect Term Extraction





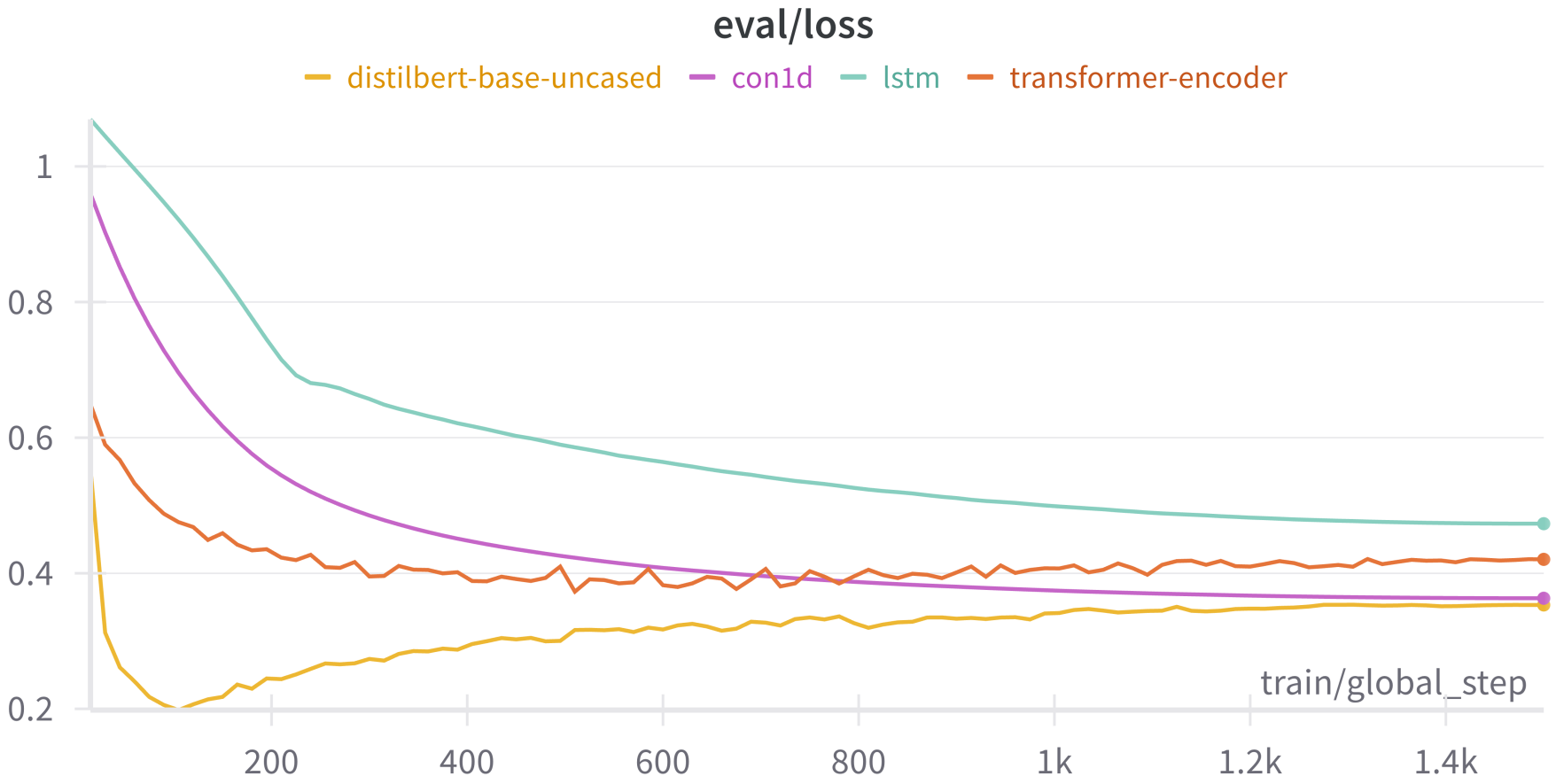
Aspect Term Extraction

Comparison



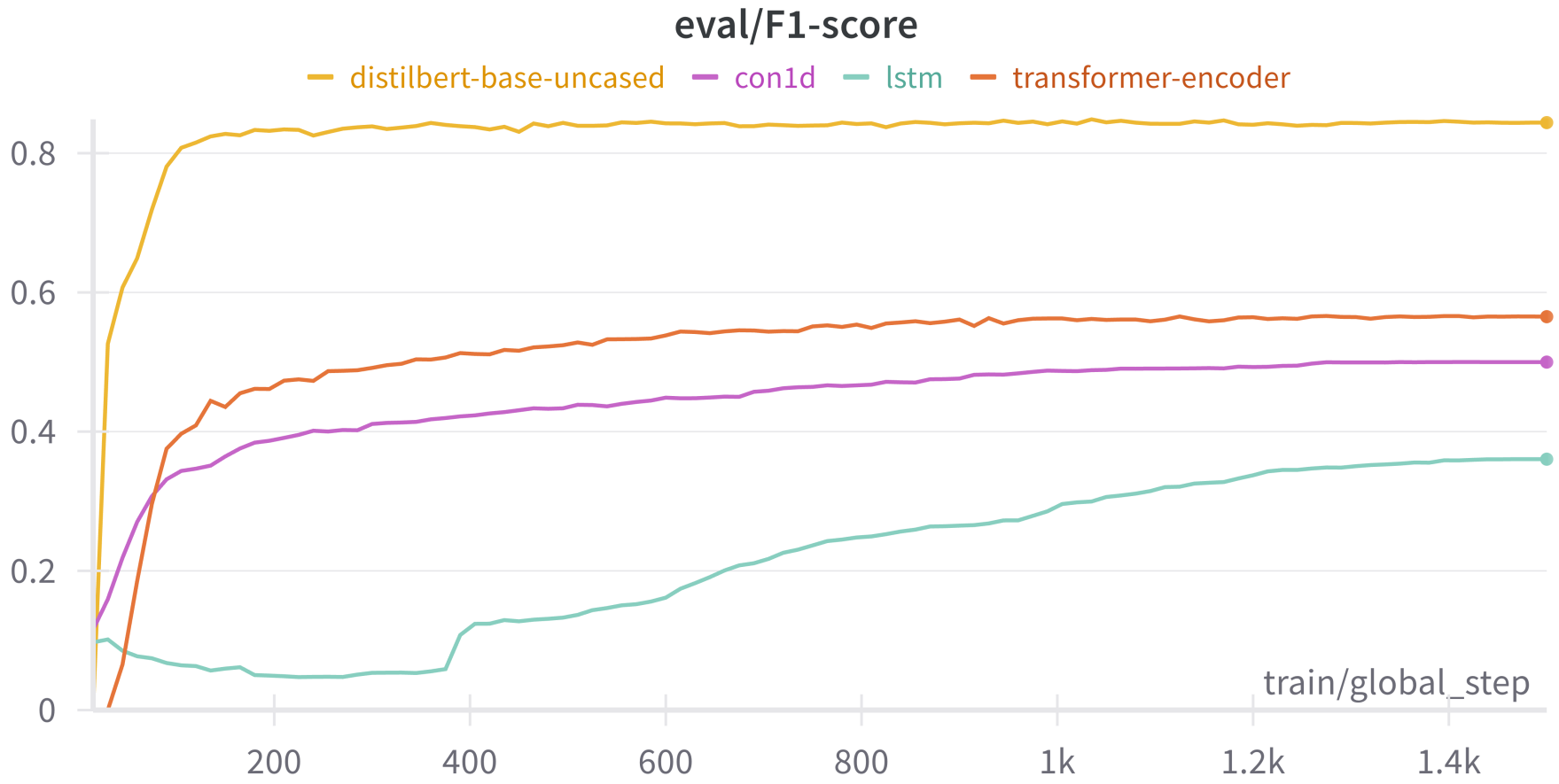
Aspect Term Extraction

Comparison



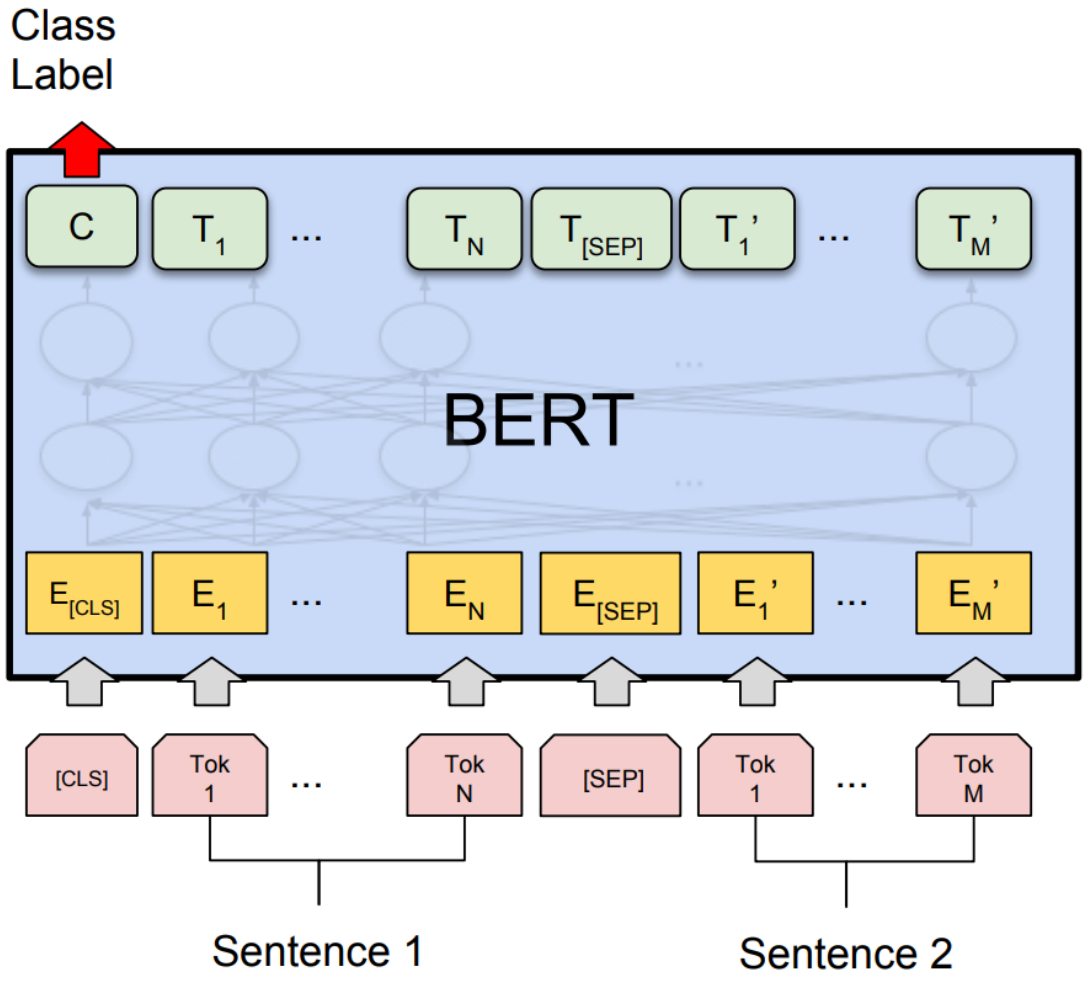
Aspect Term Extraction

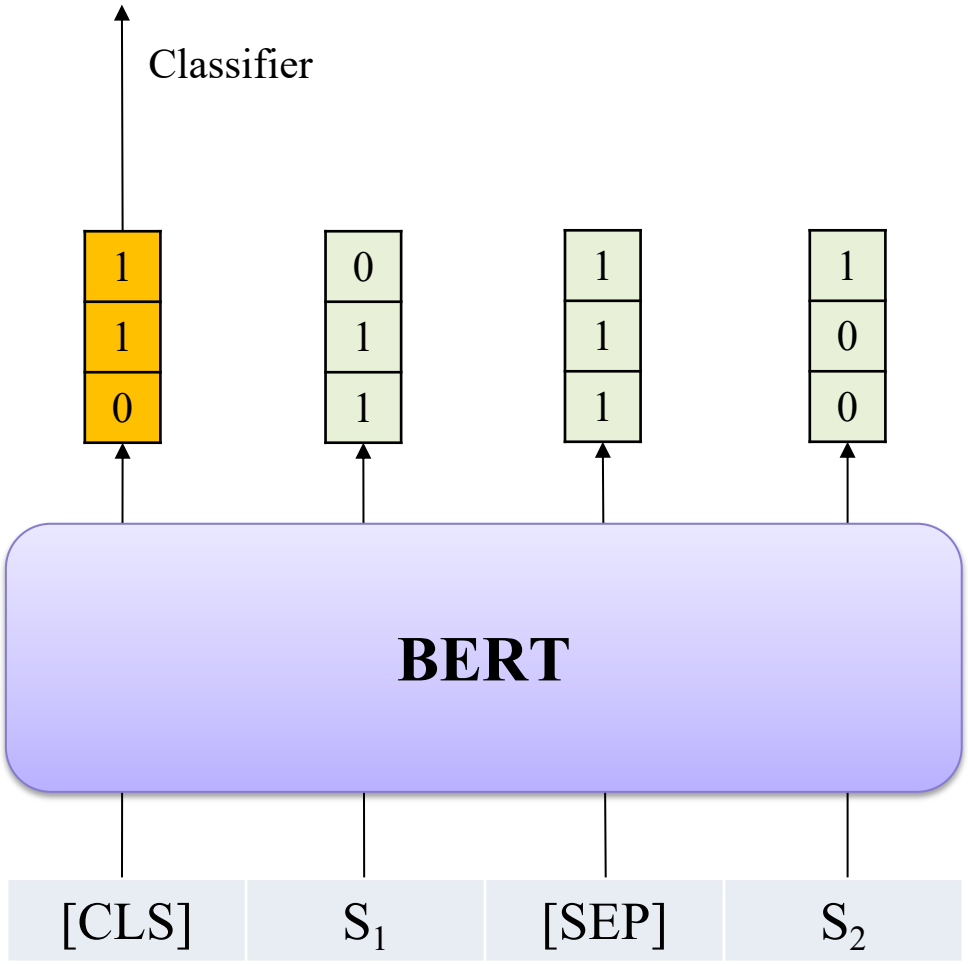
Comparison



Aspect Sentiment Pair Extraction

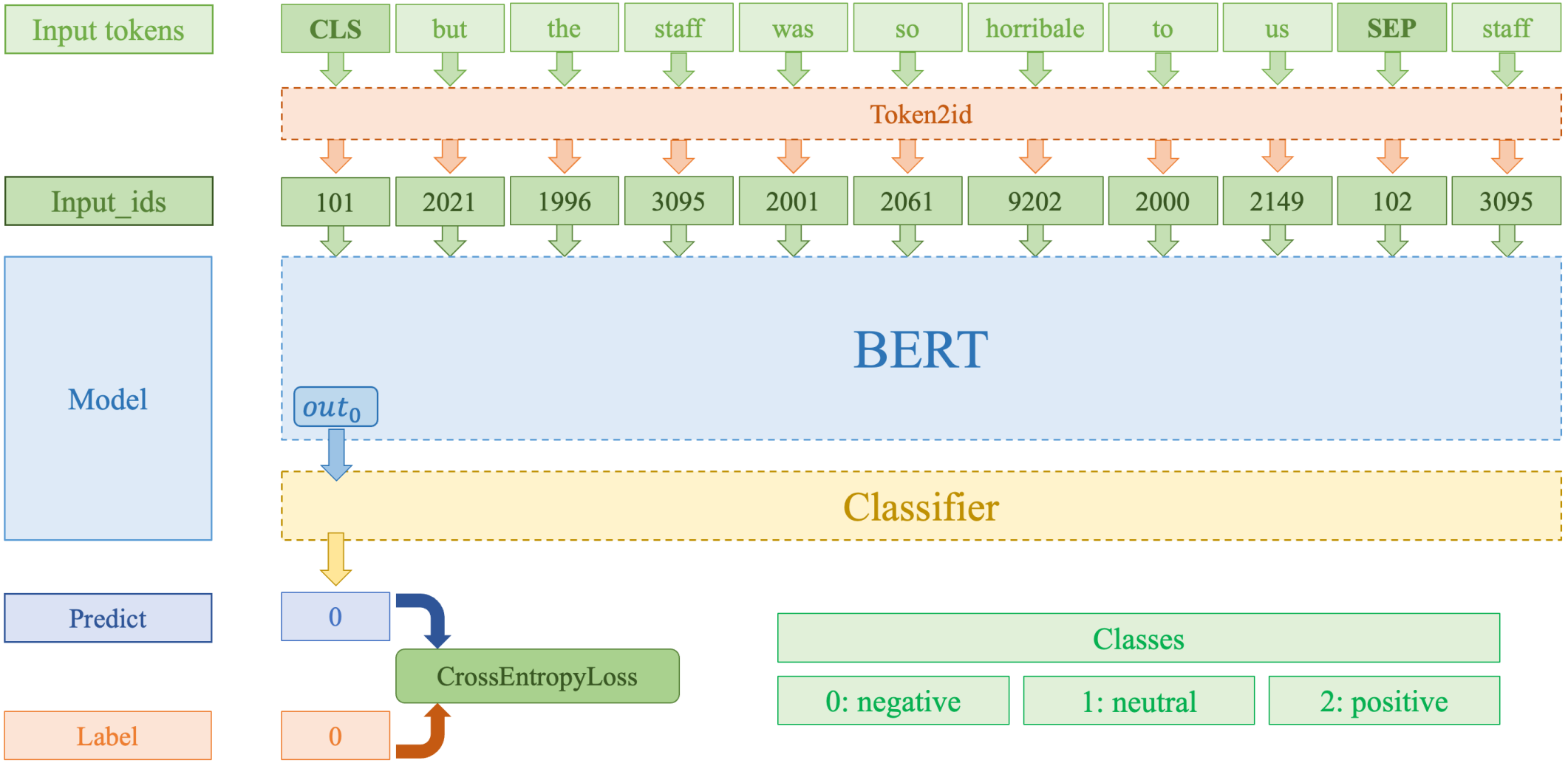
Aspect Term Sentiment Extraction (Next Sentence Prediction Task)





Aspect Sentiment Pair Extraction

Aspect Term Sentiment Extraction (Next Sentence Prediction Task)



Aspect Sentiment Pair Extraction

Doc	Label
The **staff** was so horrible	[0, **1**, 0, 0, 0, 0, 0, 0]
The **bread** is top notch as well	[0, **1**, 0, 0, 0, 0, 0, 0]

TERM

Negative - 0 Neutral - 1 Positive - 2

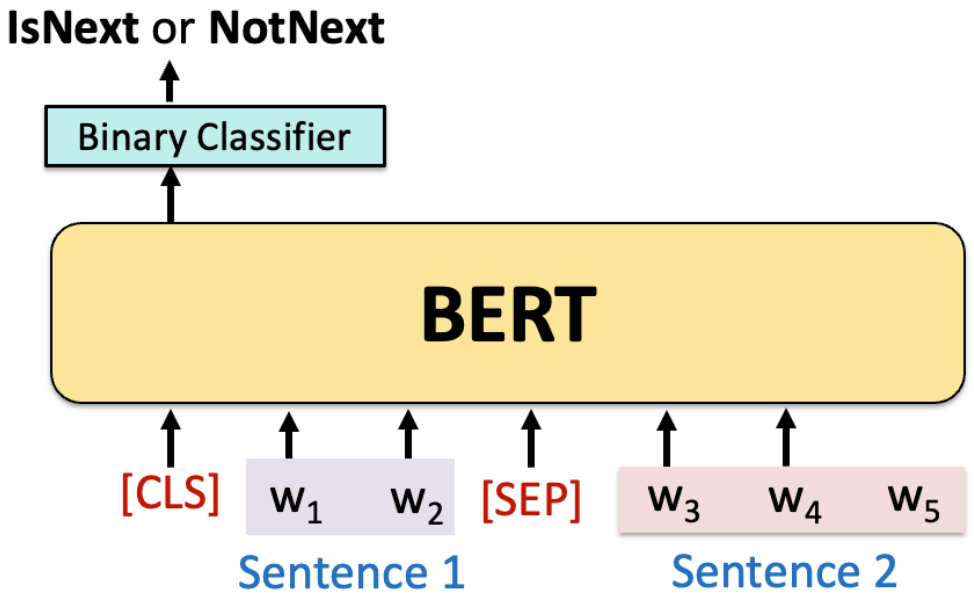
SENTIMENT

Doc	Term	Label
The **staff** was so horrible	**staff**	0 Negative
The **bread** is top notch as well	**bread**	2 Positive

Input 1 Input 2

Aspect Sentiment Pair Extraction

Next Sentence Prediction



Negative - 0 Neutral - 1 Positive - 2

SENTIMENT

Doc	Term	Label
The **staff** was so horrible	**staff**	0 Negative
The **bread** is top notch as well	**bread**	2 Positive

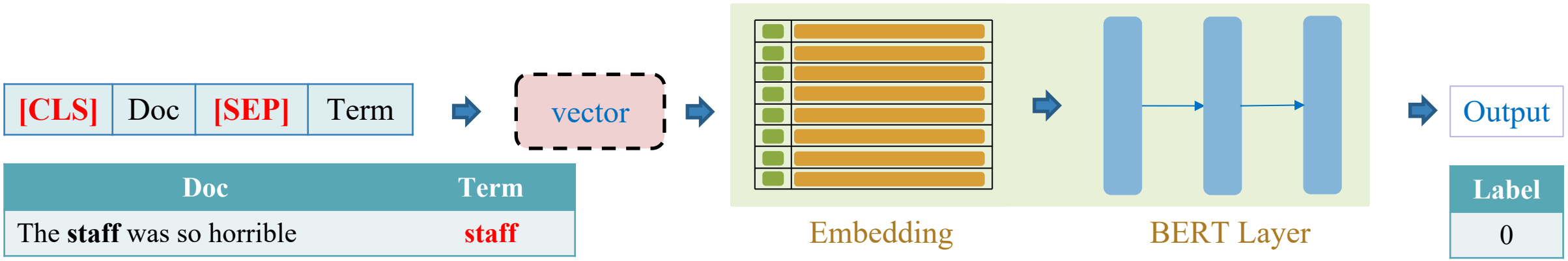
Input 1 Input 2

Aspect Sentiment Pair Extraction

Negative - 0 Neutral - 1 Positive - 2

SENTIMENT

Doc	Term	Label
The **staff** was so horrible	**staff**	0
The **bread** is top notch as well	**bread**	2



Doc	Term
The **staff** was so horrible	**staff**

Aspect Sentiment Pair Extraction

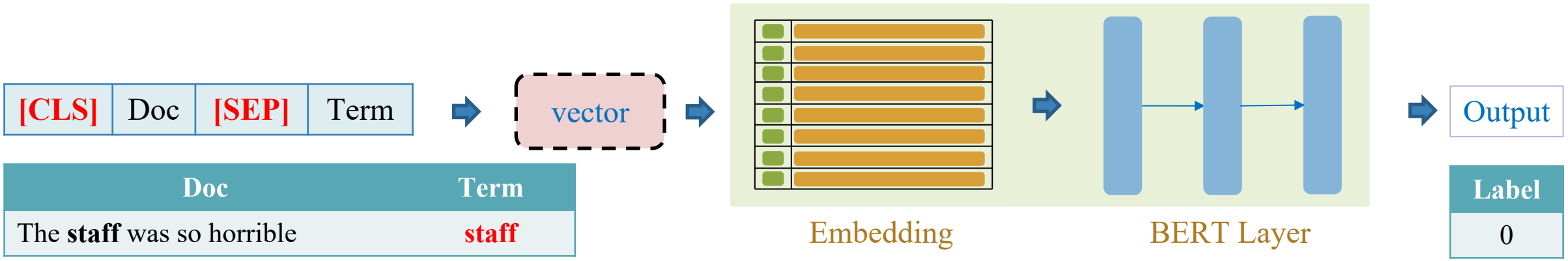
Baseline Models

RNNs Conv1D

Transformer-Encoder

Pretrained Models

BERTs



Doc	Term
The **staff** was so horrible	**staff**

Label
0

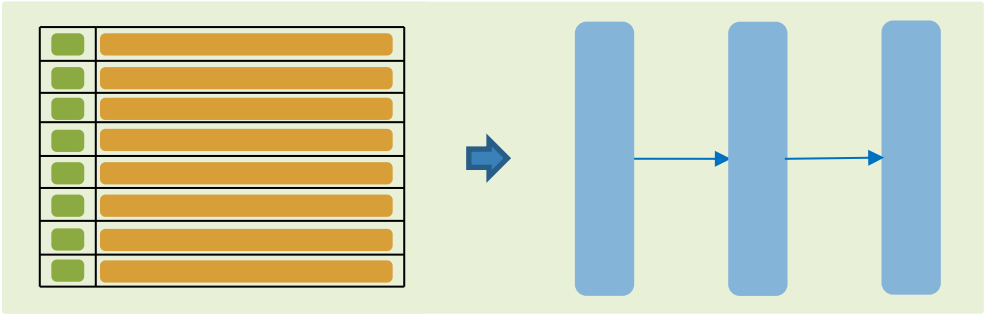
Aspect Sentiment Pair Extraction

index	token
0	[UNK]
1	[pad]
2	the
3	staff
4	was
5	so
6	horrible
…	…
30522	…

sample 1	vec 1	token_type 1	label 1
[CLS]	10	0	0
the	2	0	
staff	3	0	
was	4	0	
so	5	0	
horrible	6	0	
[SEP]	20	1	
staff	6	1	

[CLS]	Doc	[SEP]	Term

vector



Embedding BERT Layer

Output

Doc	Term
The **staff** was so horrible	staff

Label
0

BERT

Aspect Sentiment Pair Extraction

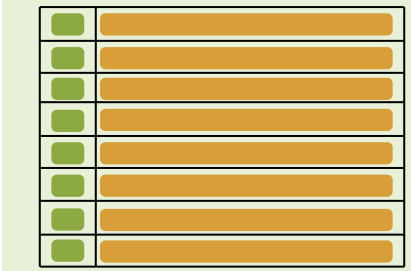
index	token
0	[UNK]
1	[pad]
2	the
3	staff
4	was
5	so
6	horrible
…	…
30522	…

sample 1	vec 1	token_type 1	label 1
[CLS]	10	0	
the	2	0	
staff	3	0	
was	4	0	
so	5	0	0
horrible	6	0	
[SEP]	20	1	
staff	6	1	

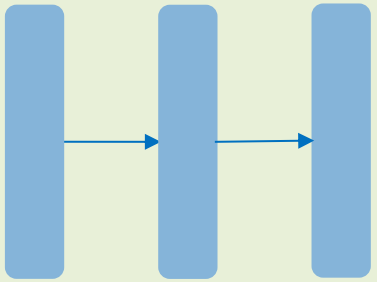
```
bert_tokens = []
bert_att = []
pols_label = 0
for i in range(len(tokens)):
    t = tokenizer.tokenize(tokens[i])
    bert_tokens += t
    if int(pols[i]) != -1:
        bert_att += t
        pols_label = int(pols[i])
```

[CLS]	Doc	[SEP]	Term

vector



Embedding



BERT Layer

Output

Doc	Term
The **staff** was so horrible	staff

Label
0

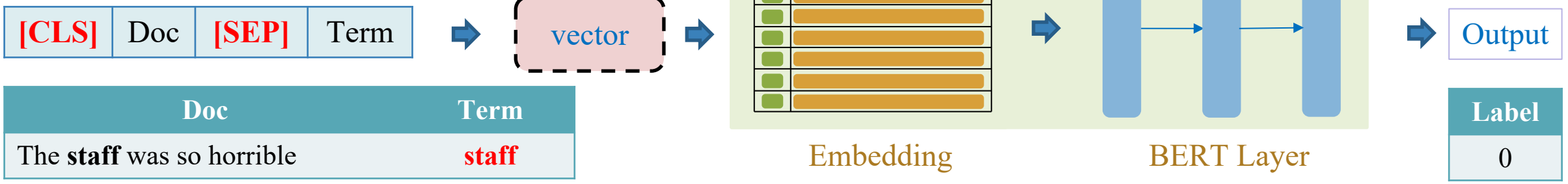
BERT

Aspect Sentiment Pair Extraction

[CLS]	10	0	[0.0445, 0.0600, 0.0663, 0.0871]
the	2	0	[-0.0266, 0.1208, 0.0685, 0.1349]
staff	3	0	[-0.0990, -0.0203, -0.0767, 0.0217]
was	4	0	[0.0909, 0.0729, -0.0171, 0.0400]
so	5	0	[0.0135, -0.0387, -0.0450, 0.1164]
horrible	6	0	[-0.0485, -0.0738, 0.1125, 0.0744]
[SEP]	20	1	[-0.1633, -0.0327, 0.0422, -0.0897]
staff	6	1	[0.0475, -0.1148, -0.1287, -0.0156]
sample 1	vec 1	token_type 1	

0

label 1

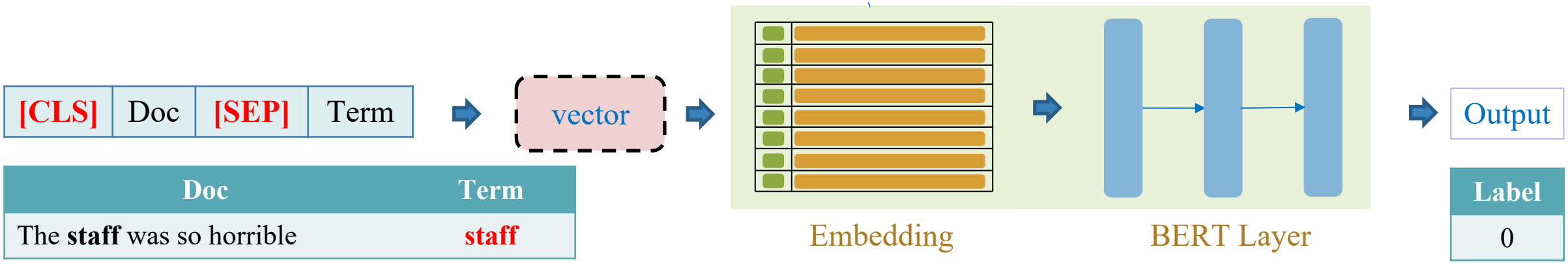


Doc	Term
The **staff** was so horrible	staff

Label
0

Aspect Sentiment Pair Extraction

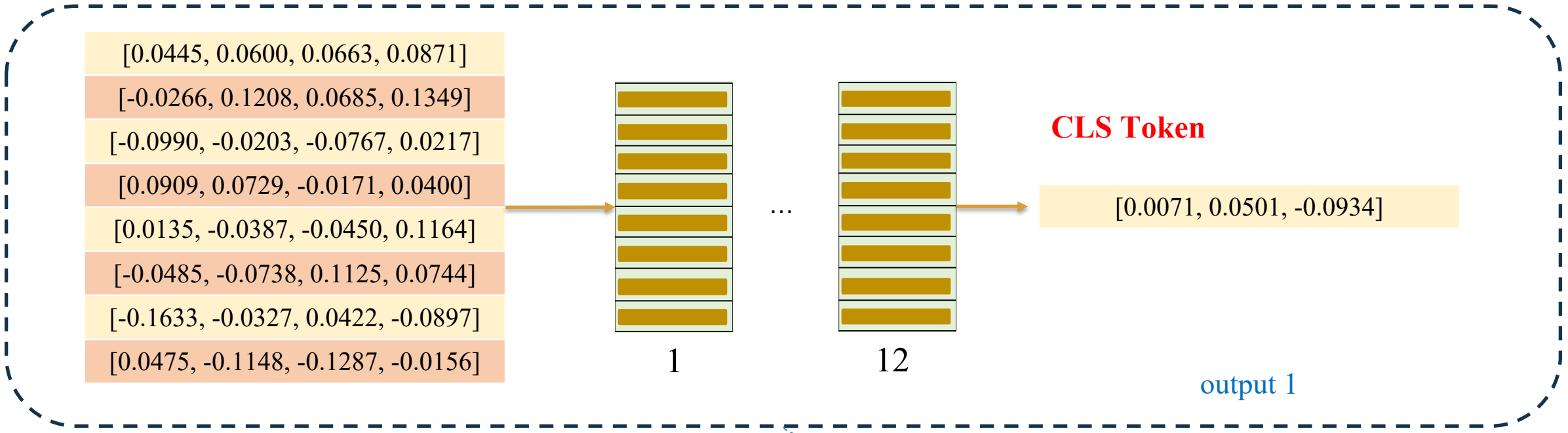
```python
from transformers import AutoModelForSequenceClassification

id2label = {0: 'Negative', 1: 'Neutral', 2: 'Positive'}
label2id = {'Negative': 0, 'Neutral': 1, 'Positive': 2}

model = AutoModelForSequenceClassification.from_pretrained(
    "distilbert/distilbert-base-uncased",
    num_labels=3, id2label=id2label, label2id=label2id
)
```

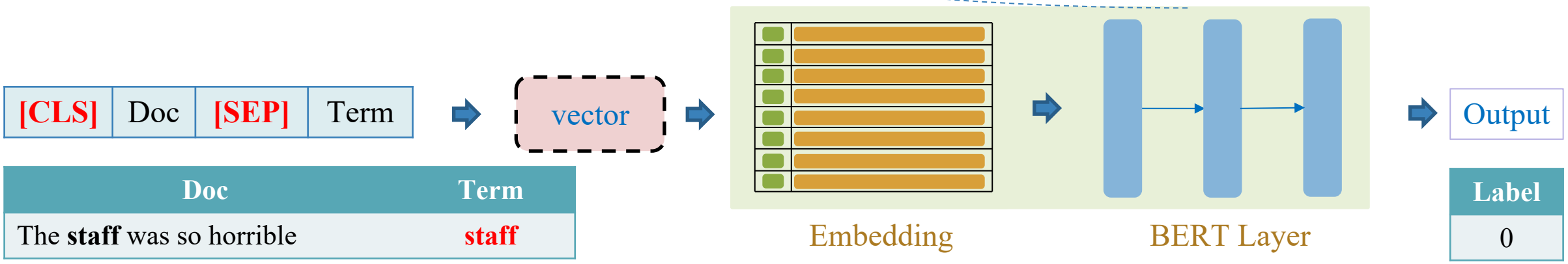


Doc	Term
The **staff** was so horrible	staff

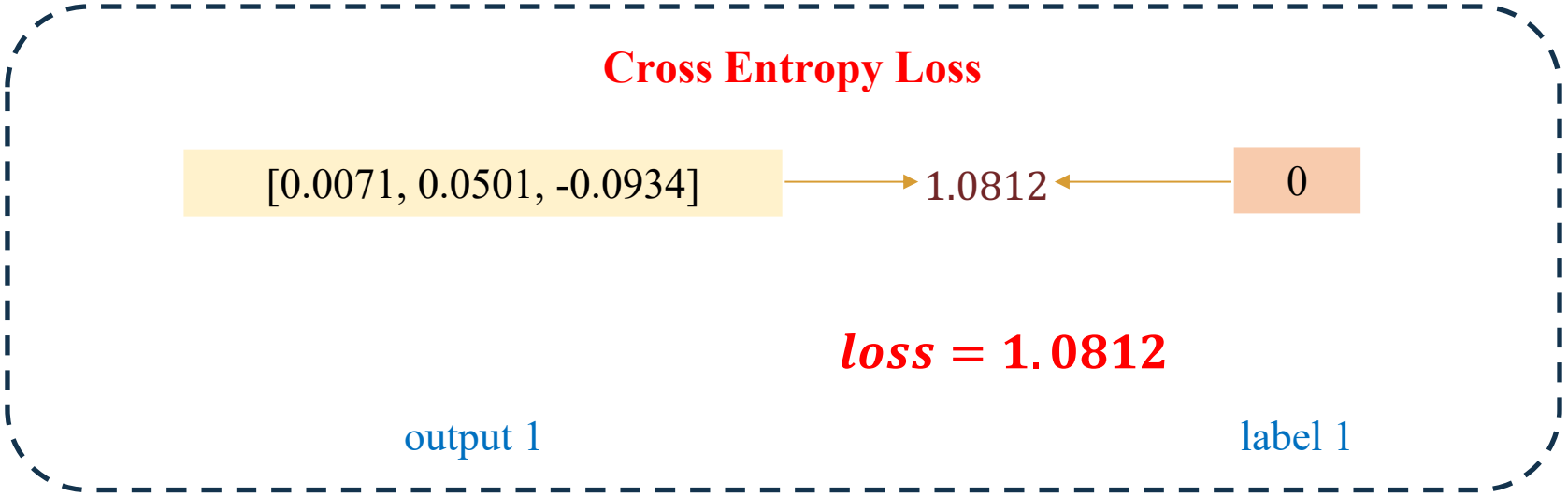
Label
0

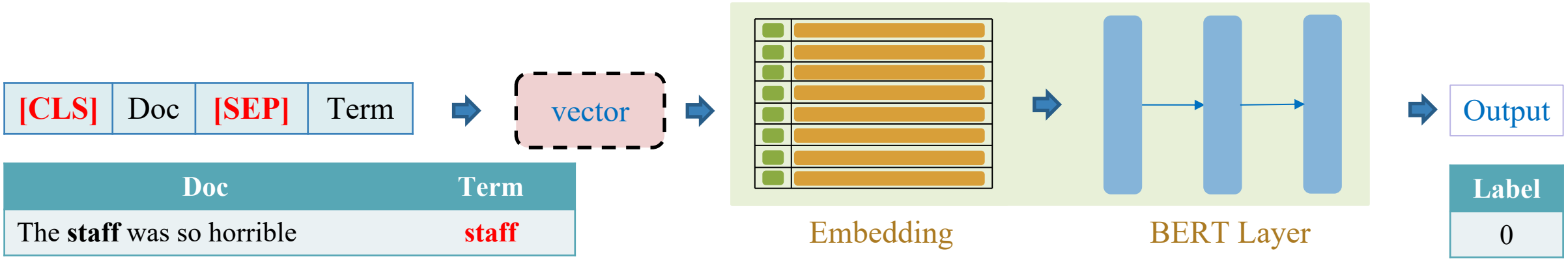
Aspect Sentiment Pair Extraction





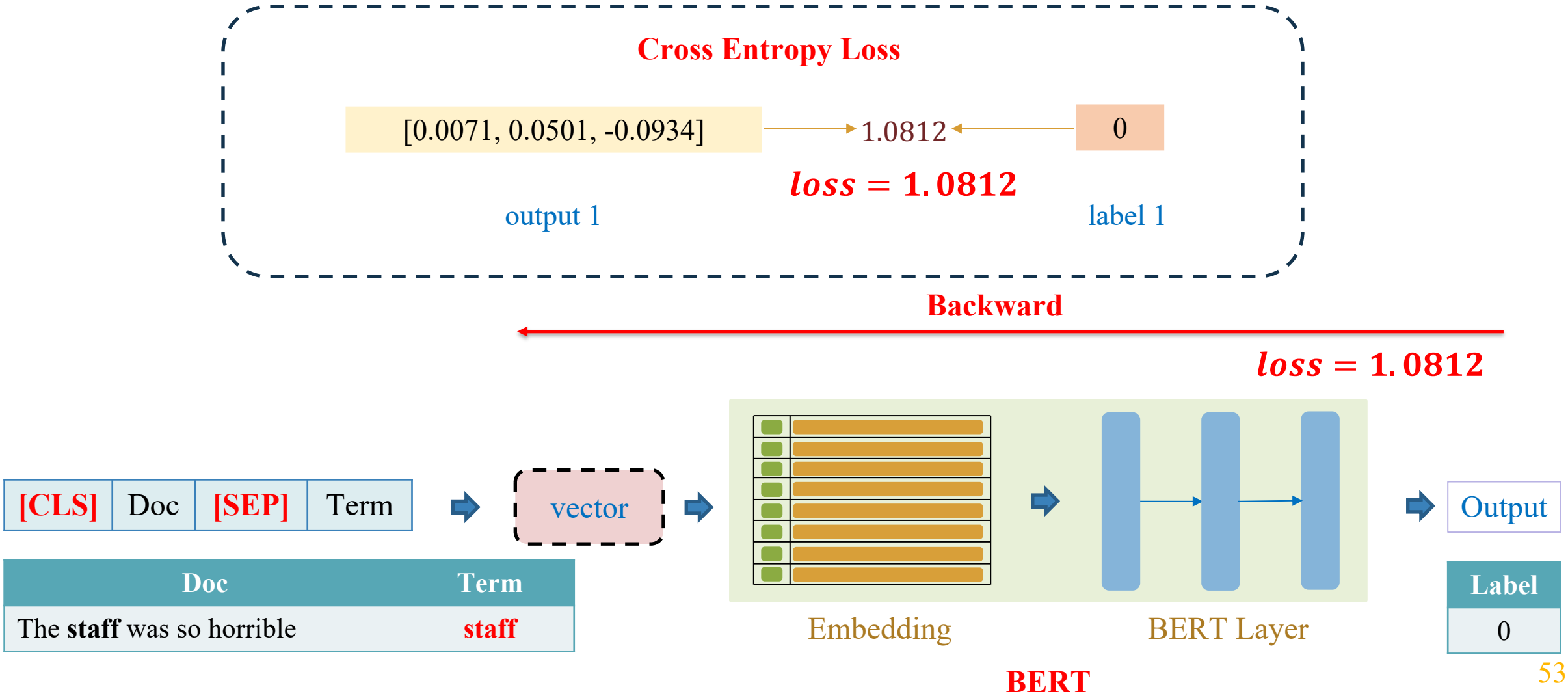
Aspect Sentiment Pair Extraction





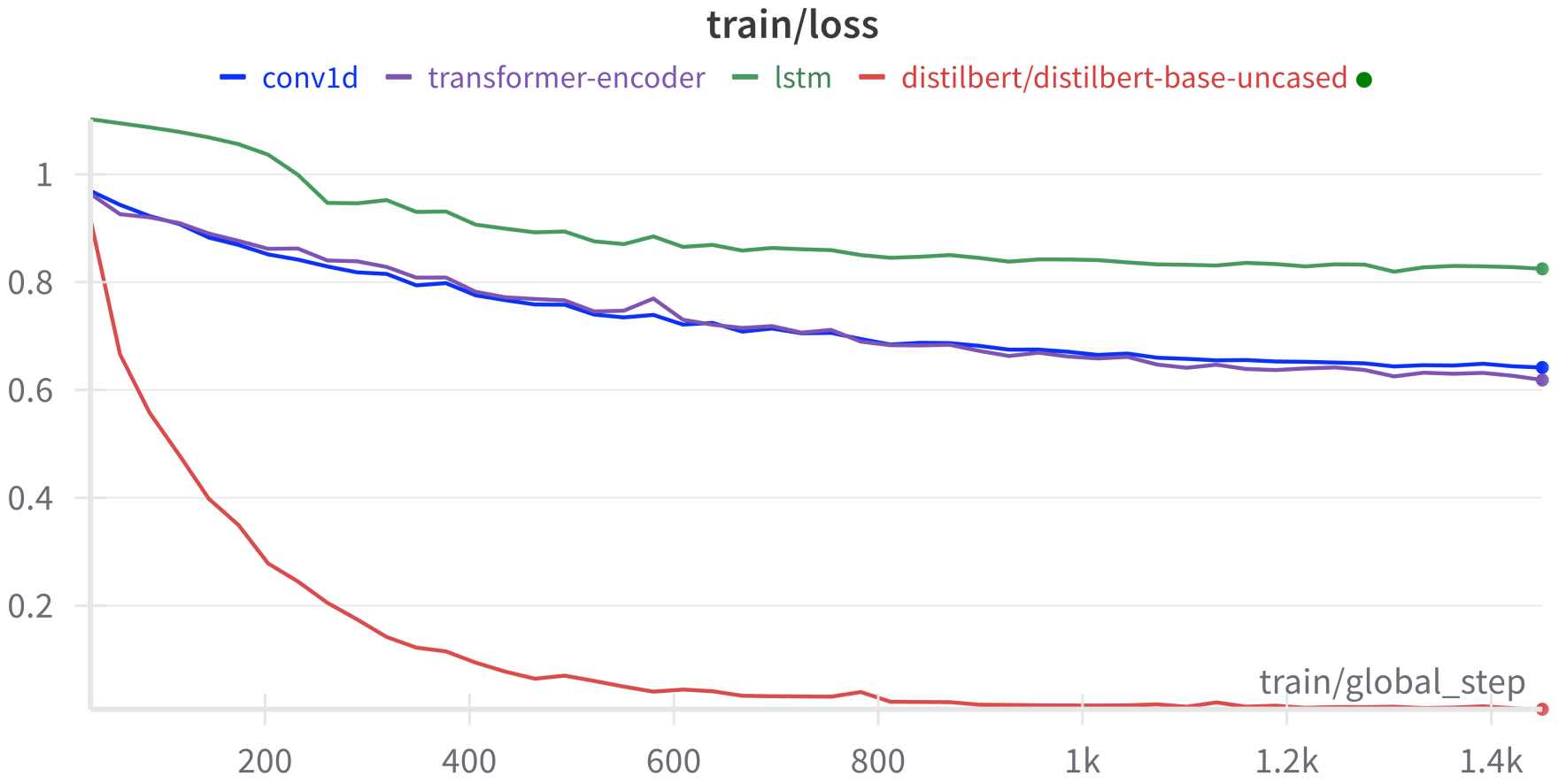
Doc	Term
The **staff** was so horrible	staff

Aspect Sentiment Pair Extraction



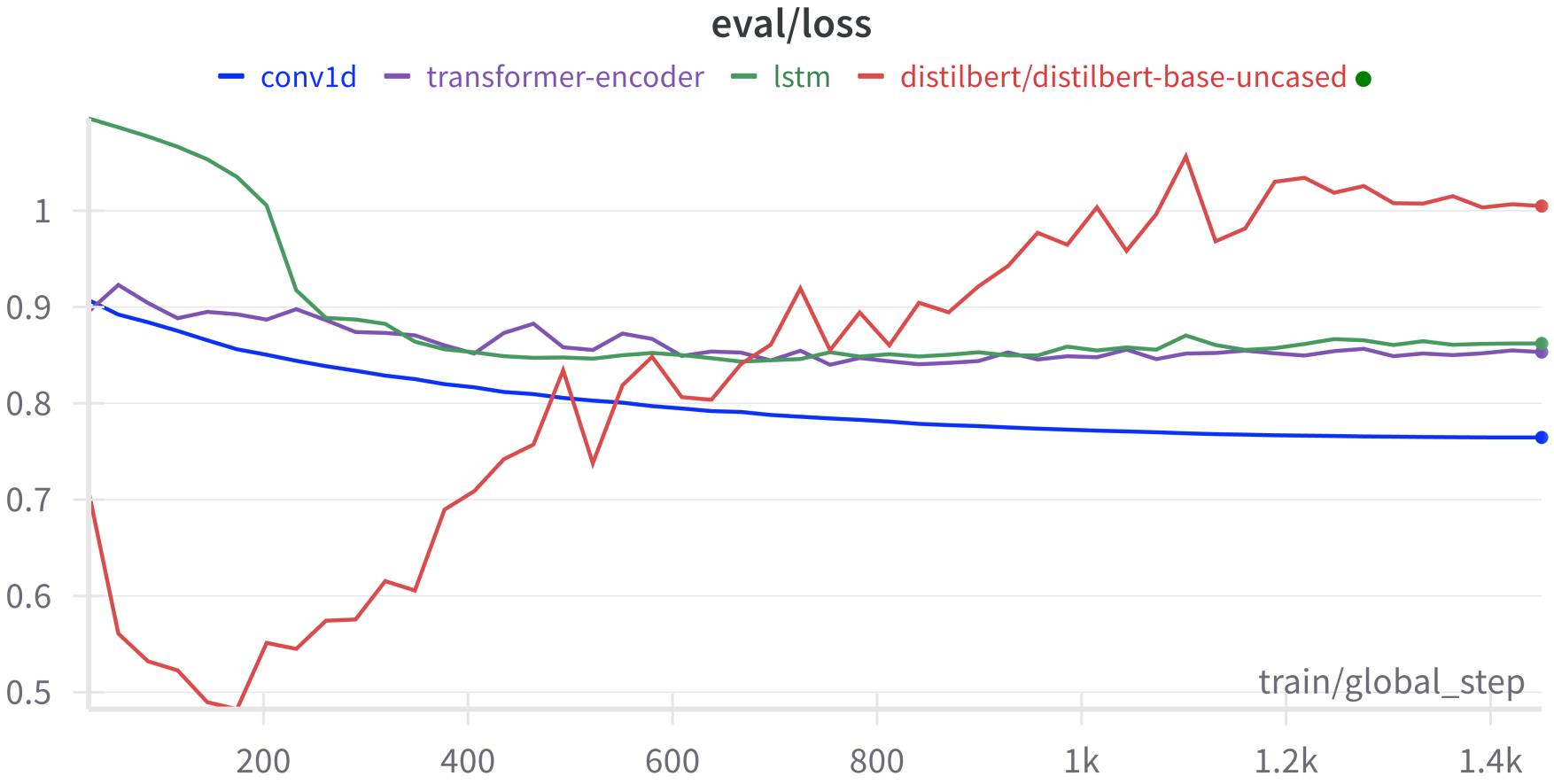
Aspect Sentiment Pair Extraction

Comparison



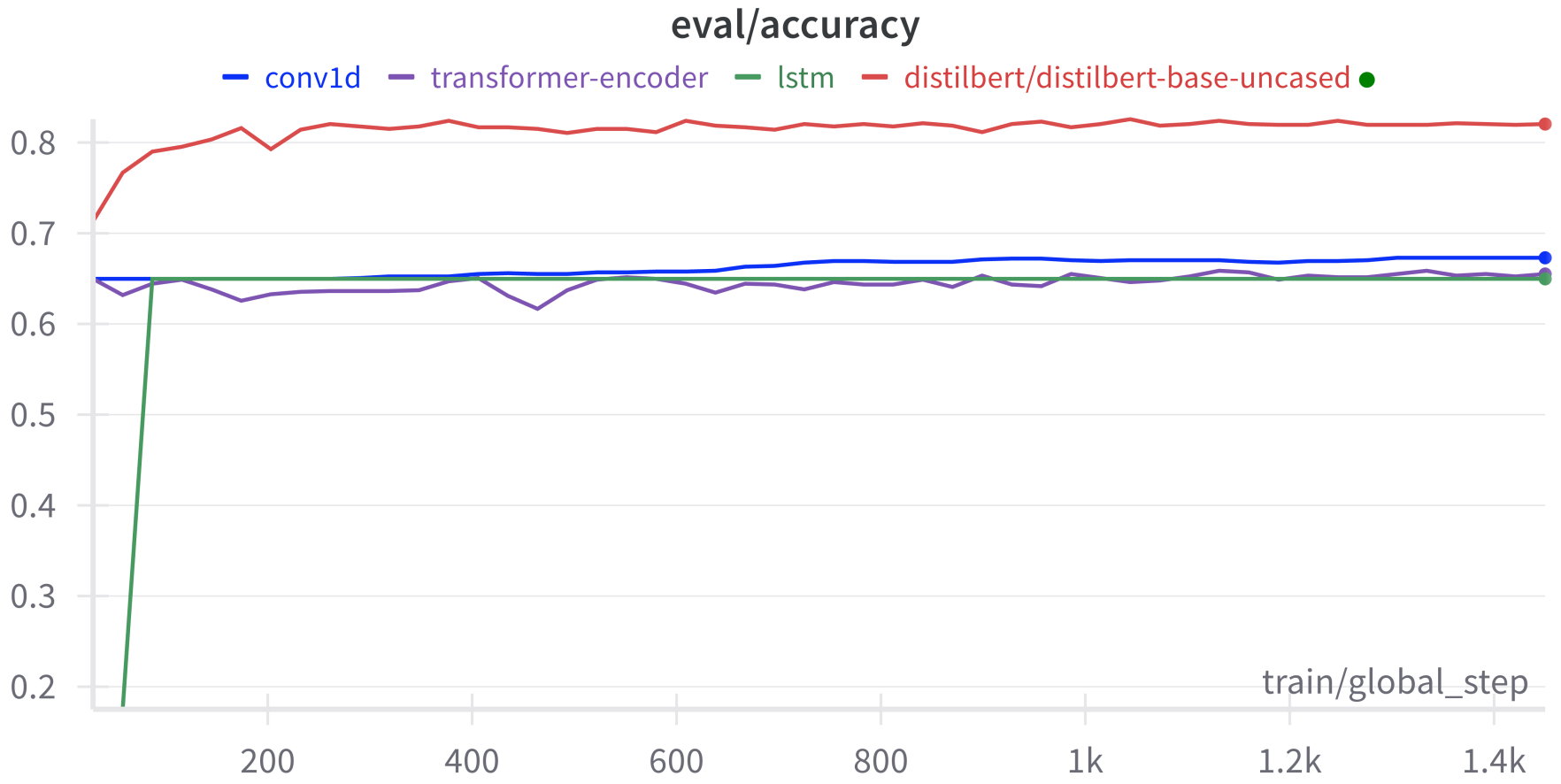
Aspect Sentiment Pair Extraction

Comparison



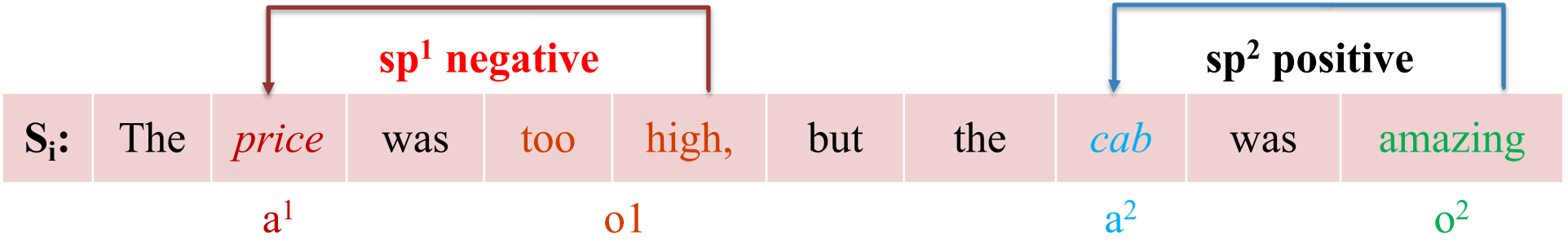
Aspect Sentiment Pair Extraction

Comparison

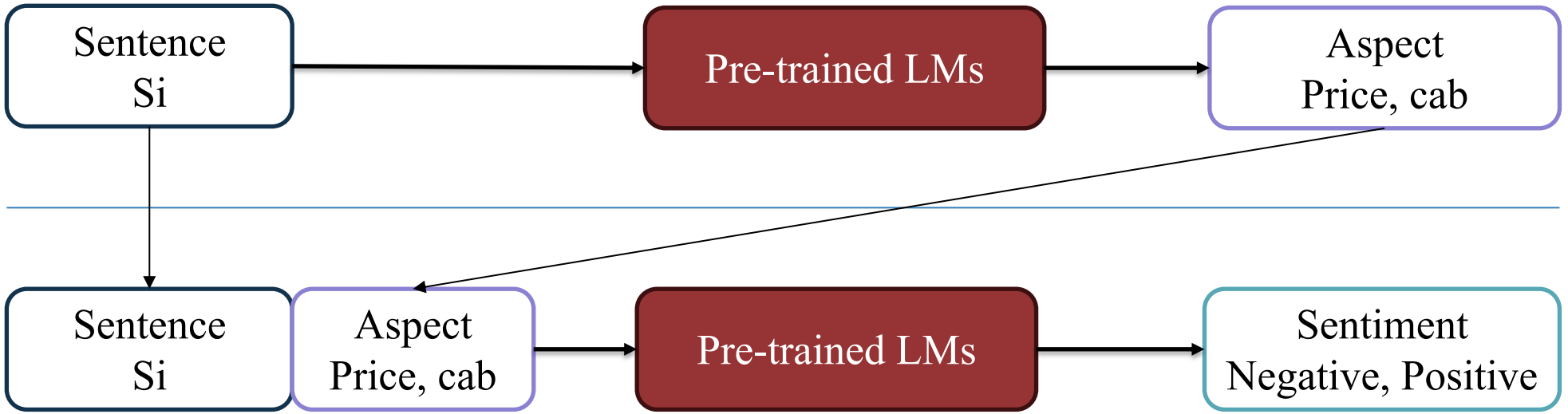


Aspect Sentiment Pair Extraction

Summary



Stage 1: Aspect Term Extraction (Token-level Classification)



Stage 2: Aspect Term Sentiment Extraction (Document-level Classification)

Aspect Sentiment Pair Extraction

Summary

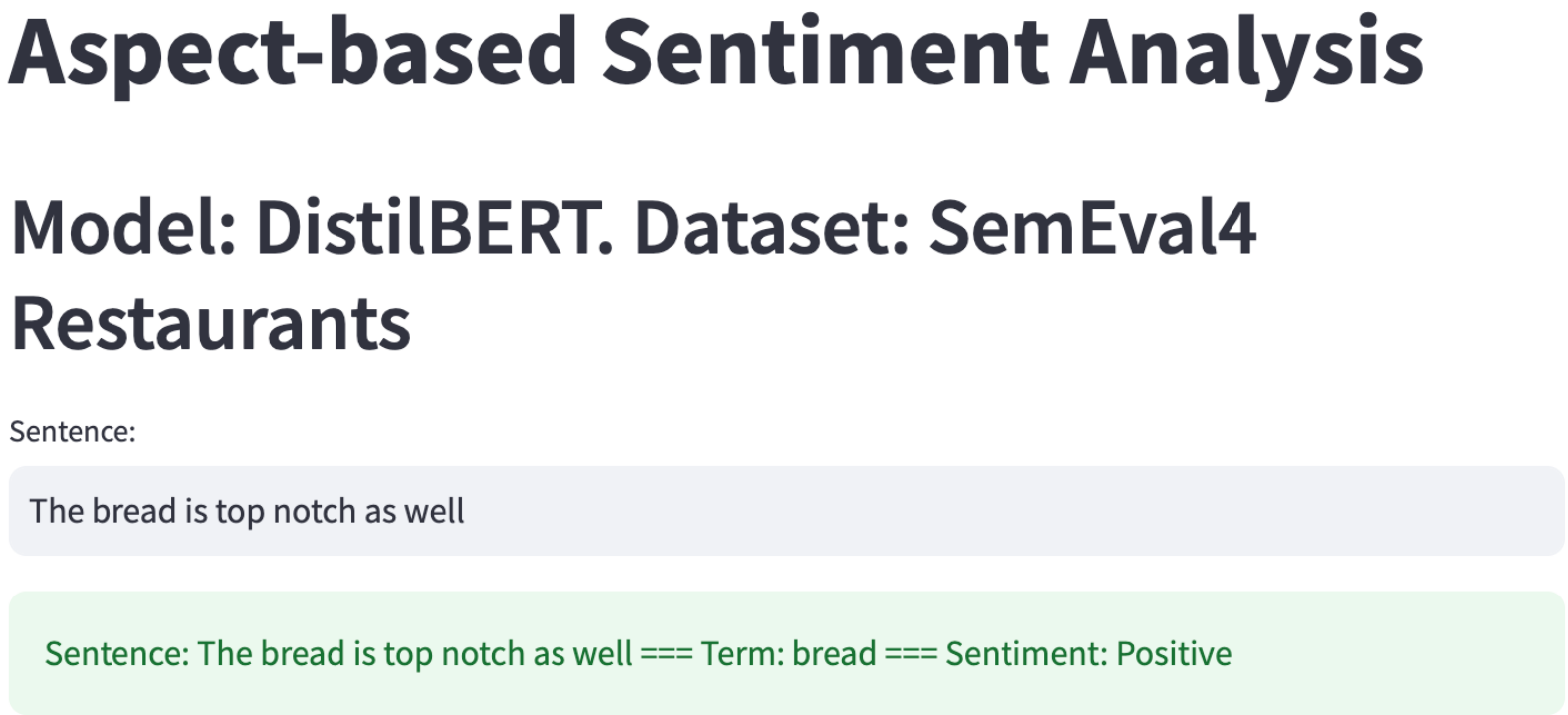
```python
from transformers import pipeline

token_classifier = pipeline(
    model="thainq107/abte-restaurants-distilbert-base-uncased",
    aggregation_strategy="simple"
)

classifier = pipeline(
    model="thainq107/absa-restaurants-distilbert-base-uncased"
)
```

```python
test_sentence = 'The bread is top notch as well'
results = token_classifier(test_sentence)
sentence_tags = " ".join([result['word'] for result in results])
pred_label = classifier(f'{test_sentence} [SEP] {sentence_tags}')
sentence_tags, pred_label
```

```
('bread', [{'label': 'Positive', 'score': 0.9864555597305298}])
```

Aspect Sentiment Pair Extraction

Deployment (Github – Demo)




QUIZ TIME

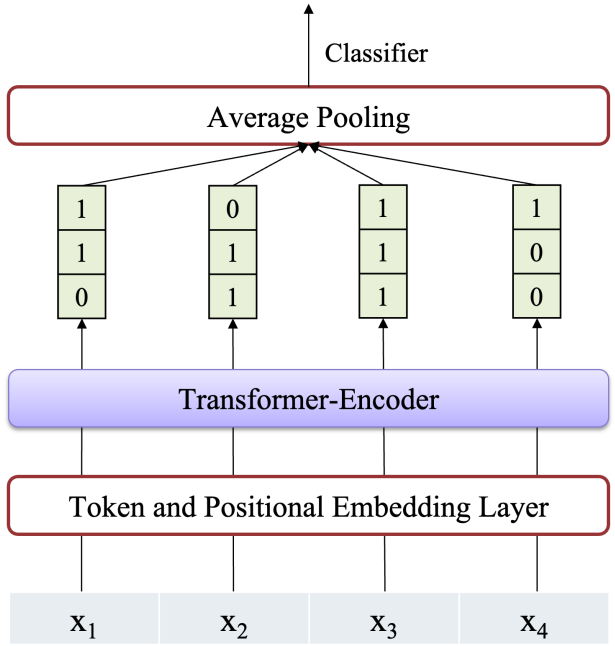
Objectives

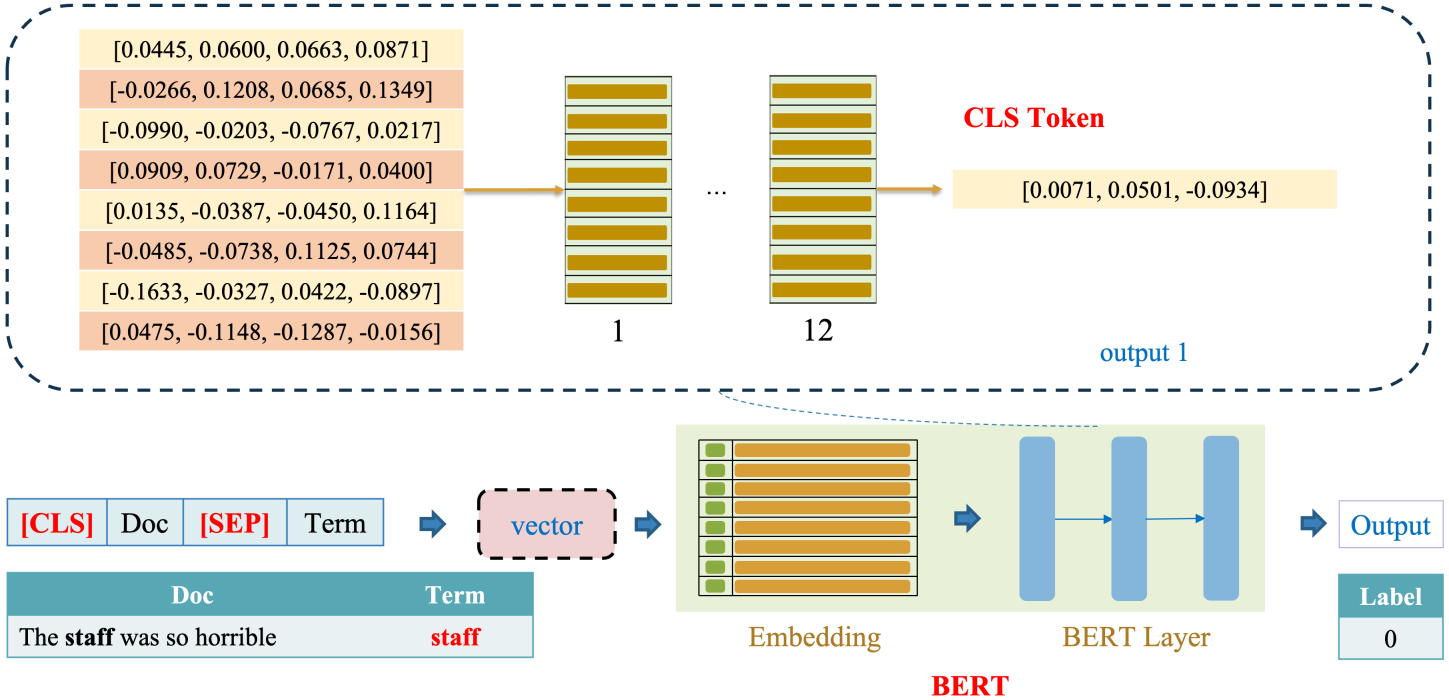
Sentiment Analysis

- Text Classification
- Types of Sentiment Analysis
- Classifier

Aspect-based Sentiment Analysis

- Subtasks
- Aspect Term Extraction
- Aspect Sentiment Pair Extraction





Thanks!

Any questions?