復旦大學 FUDAN UNIVERSITY

FNLP The Fudan Lab For NLP

Does syntax matter? A strong baseline for Aspect-based Sentiment Analysis with RoBERTa

Junqi Dai*, Hang Yan*, Tianxiang Sun, Pengfei Liu, Xipeng Qiu

jqdai19@fudan.edu.cn

NAACL 2021

Outline

ALSC models and tree structure

PTMs and tree structure

Experiments: different trees on different ALSC models

Analysis: Discussions and Surprise

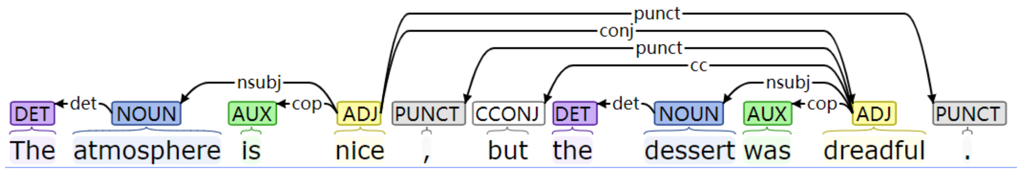
Conclusion

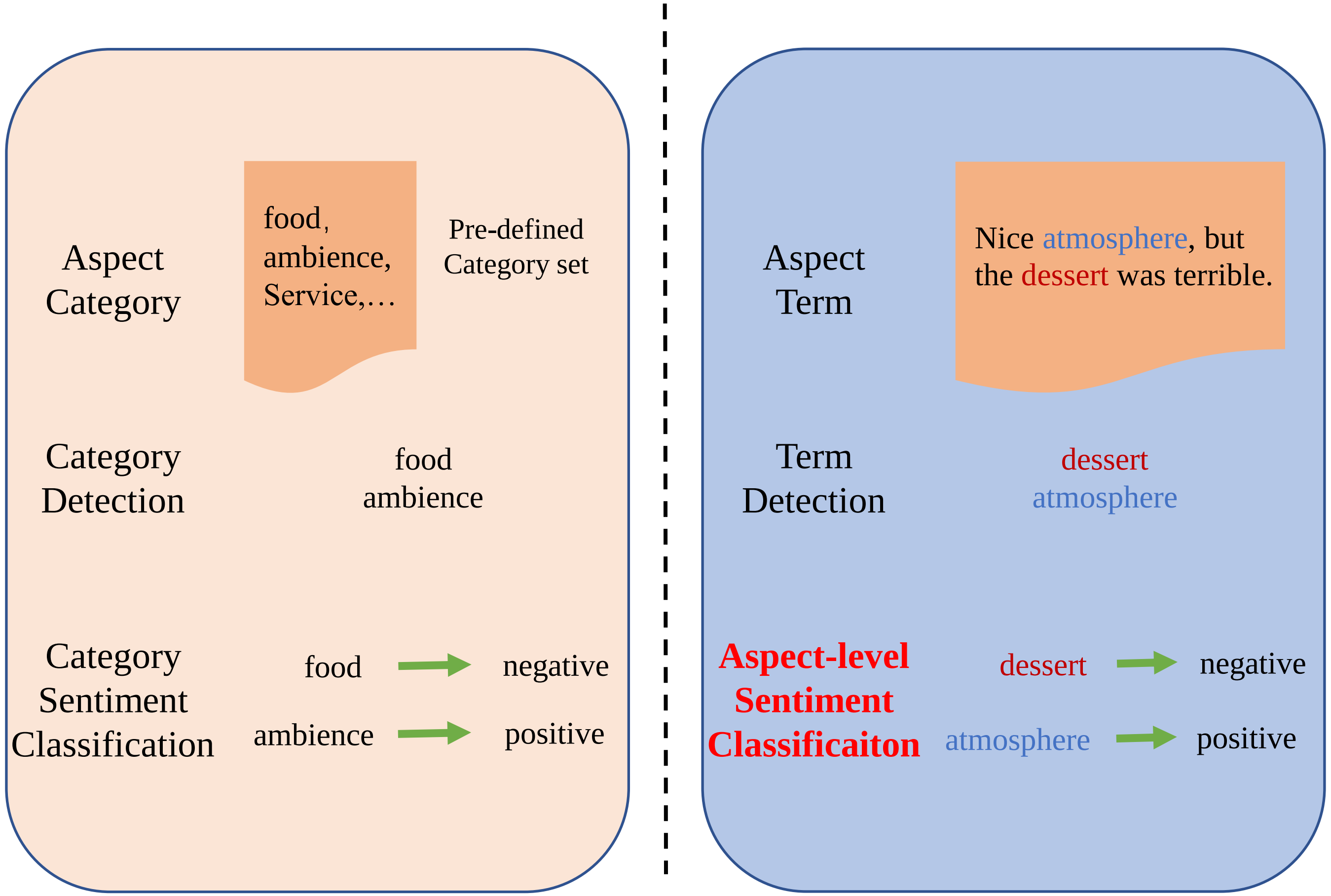
What is the ALSC?

FUDAN UNIVERSITY 復旦大學

The atmosphere is nice , but the dessert was dreadful.

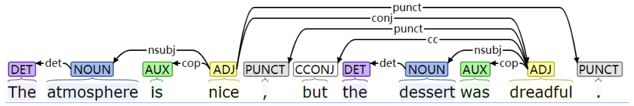
The atmosphere is nice , but the dessert was dreadful.			
Aspect Category		**Aspect Term**	
food	ambience	dessert	atmosphere
negative	positive	negative	positive





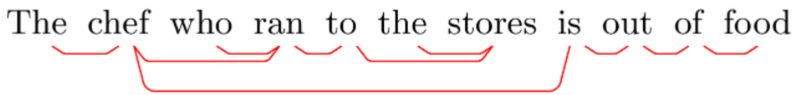
Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In Association for Computational Linguistics (ACL) System Demonstrations. 2020.

ALSC models and tree structure

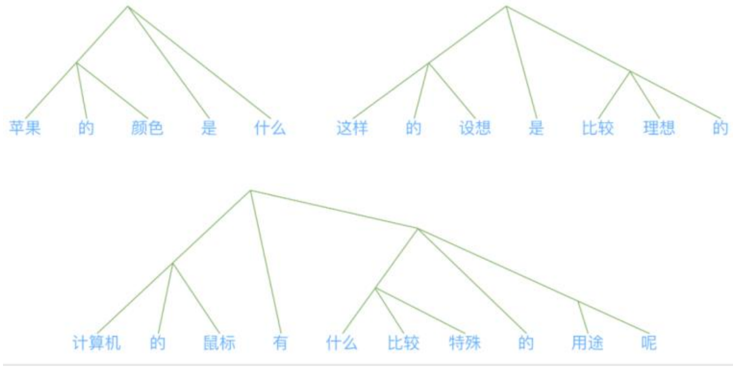
Tree	Implementation method		Selected model
	Topological Structure	**dependent relation** dessert <-- “the” “dreadful” Atmosphere <-- “nice” “,” “dreadful” …	ASGCN
 Dependency result	Structure Distance	**dependent relation & distance value** dessert: “the” (1) “dreadful” (1) “nice” (3) …	RGAT
	Tree-based Distance	**distance value** dessert： 4 3 3 2 2 2 1 0 2 1 3 …	PWCN

ASGCN: Aspect-Level Sentiment Analysis Via Convolution over Dependency Tree, Sun et al. (2019)
RGAT: Relational Graph Attention Network for Aspect-based Sentiment Analysis, Wang et al. (2020)
PWCN: Syntax-Aware Aspect-Level Sentiment Classification with Proximity-Weighted Convolution Network, Zhang et al. (2019)

PTMs and tree structure



Structural Probe



Structural Probe on Chinese

PTMs

↓

$$f(x_i,x_j)=||H_\theta(\mathbf{x}\backslash\{x_i\})_i-H_\theta(\mathbf{x}\backslash\{x_i,x_j\})_i||_2$$

Impact Matrix

↓ Tree decoding algorithm

Dependency tree

Perturbed Masking Method

A Structural Probe for Finding Syntax in Word Representations, Hewitt and Manning (2019)
Perturbed Masking: Parameter-free Probing for Analyzing and Interpreting BERT, Wu et al. (2020)
苏剑林. (Jun. 10, 2020). 《无监督分词和句法分析！原来BERT还可以这样用 》[Blog post]. Retrieved from https://spaces.ac.cn/archives/7476

Questions

What about the comparison between:

1. Tree induced from PTMs vs. Tree from dependency parser ?

2. Tree induced from PTMs vs. Tree from **task fine-tuned** PTMs ?

Experiments: Experimental Setup

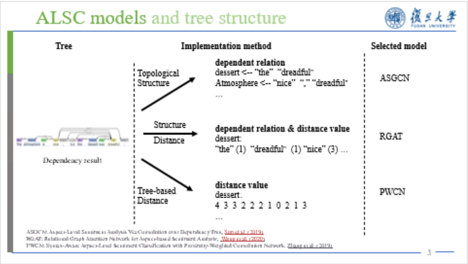
- Task: Aspect-level Sentiment Classification
- Datasets:

  Rest14 Pontiki et al., (2014)

  Laptop14 Pontiki et al., (2014)

  Twitter Dong et al. (2014)

- ALSC models



- PTMs: BERT, RoBERTa

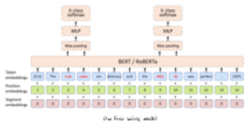
Experiments: Trees

Pre-trained BERT → Perturbed Masking → **BERT Induced Tree**

Pre-trained RoBERTa → Perturbed Masking → **RoBERTa Induced Tree**

Pre-trained BERT → Fine-tune → Perturbed Masking → **FT-BERT Induced Tree**

Pre-trained RoBERTa → Fine-tune → Perturbed Masking → **FT-RoBERTa Induced Tree**



Trees

- Dep. (Dependency tree)
- Left/Right chain
- BERT Induced Tree
- RoBERTa Induced Tree
- FT-BERT Induced Tree
- FT-RoBERTa Induced Tree

Experiments: Models and Trees

ALSC models

- ASGCN
  - Topological Structure
- PWCN
  - Relative Distance
- RGAT
  - Structure & Distance

Main Experiments

Incorporate all trees with all ALSC models.

Trees

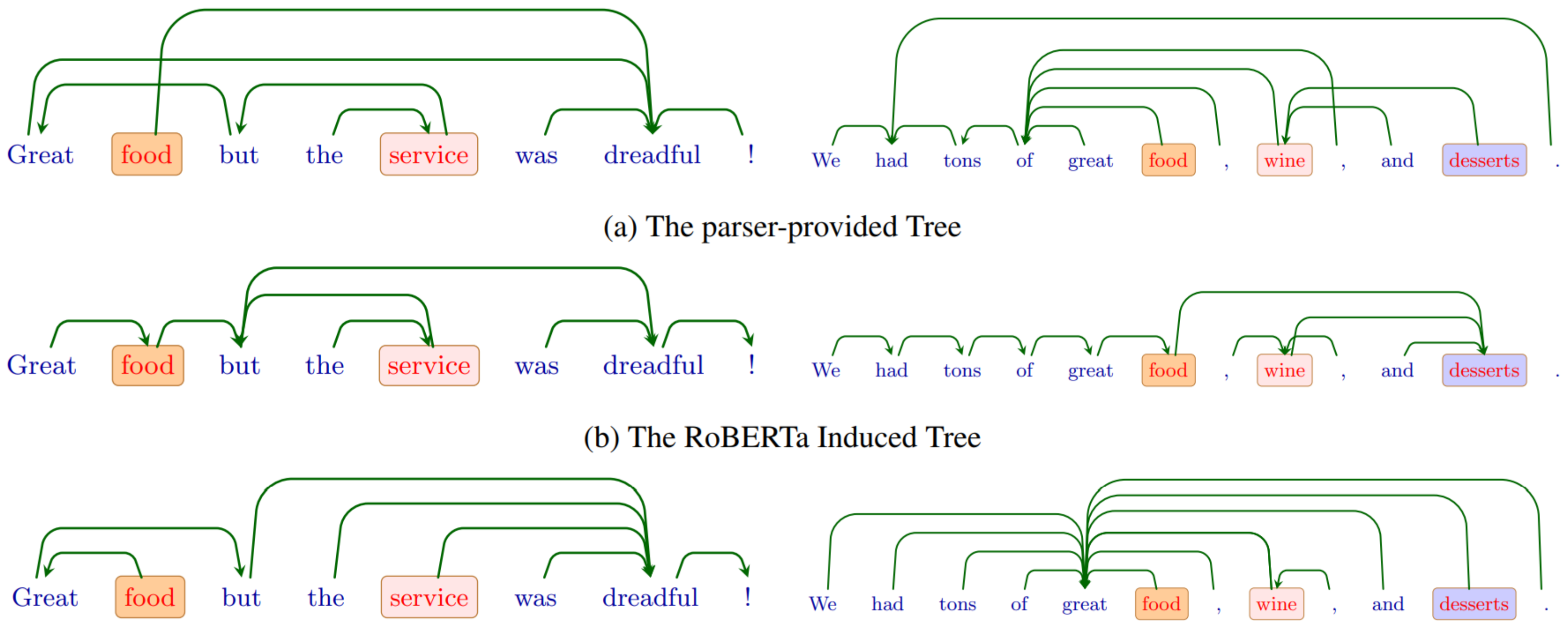
- Dep. (Dependency tree)
- Left/Right chain
- BERT Induced Tree
- RoBERTa Induced Tree
- FT-BERT Induced Tree
- FT-RoBERTa Induced Tree

Experiments: Main Results

Model	Tree Features	Tree Structure	Rest14		Laptop14		Twitter	
			Acc.	F_1	*Acc.*	F_1	*Acc.*	F_1
BiLSTM	-	-	77.59	67.05	70.06	64.46	71.39	69.45
ASGCN	Topological Structure	Zhang et al. (2019a)	80.86	72.19	75.55	71.05	72.15	70.40
		Dep.	81.42	72.87	75.54	71.66	72.36	70.32
		Left-chain	80.89	71.92	73.98	69.81	71.96	70.47
		Right-chain[4]	80.89	71.92	73.98	69.81	71.96	70.47
		BERT Induced Tree	81.07	72.87	74.29	70.42	72.39	70.25
		RoBERTa Induced Tree	81.16	72.33	74.76	70.0	72.76	71.17
		FT-BERT Induced Tree	81.87	72.89	74.85	70.71	73.36	71.61
		FT-RoBERTa Induced Tree	**82.31**	**73.53**	**76.33**	**72.76**	**73.84**	**72.66**
PWCN	Tree-based Distance	Zhang et al. (2019b)	80.96	72.21	76.12	72.12	-	-
		Dep.	80.89	72.16	75.86	71.94	72.10	70.75
		Left-chain	80.78	72.37	73.35	69.41	71.24	69.42
		Right-chain[4]	80.78	72.37	73.35	69.41	71.24	69.42
		BERT Induced Tree	80.98	72.04	73.82	69.35	72.10	69.90
		RoBERTa Induced Tree	81.16	73.20	73.98	69.94	72.11	70.74
		FT-BERT Induced Tree	81.33	73.57	74.96	70.93	72.54	70.75
		FT-RoBERTa Induced Tree	**82.40**	**73.69**	**76.95**	**73.21**	**73.84**	**71.43**
RGAT	Structure & Distance	Wang et al. (2020)	**83.30**	**76.08**	77.42	73.76	**75.57**	73.82
		Dep.	82.14	74.62	76.49	72.63	74.57	72.57
		Left-chain	80.53	69.63	74.14	70.04	73.41	71.99
		Right-chain[4]	80.53	69.63	74.14	70.04	73.41	71.99
		BERT Induced Tree	81.27	71.76	75.23	70.47	73.49	72.19
		RoBERTa Induced Tree	81.42	71.79	75.36	71.11	73.78	72.37
		FT-BERT Induced Tree	81.60	72.48	75.96	71.96	74.13	72.47
		FT-RoBERTa Induced Tree	82.76	75.25	**77.43**	**74.21**	75.43	**74.04**

- Left/Right Chain VS BERT/RoBERTa Induced Tree
- Dep. VS BERT/RoBERTa Induced Tree
- BERT/RoBERTa Induced Tree VS FT-BERT/RoBERTa Induced Tree

Analysis: Discussions



(a) The parser-provided Tree

(b) The RoBERTa Induced Tree

(c) The FT-RoBERTa Induced Tree

Analysis: Discussions

- **Proportion of Neighboring Connections**
  - calculate the proportion of neighboring connections in the sentence.
  - A neighboring connection links the word to its left/right neighbor word.

Tree Structure	Rest14	Laptop14	Twitter
Dep.	0.509	0.500	0.509
Left-chain	1.000	1.000	1.000
Right-chain	1.000	1.000	1.000
BERT	0.710	0.690	0.741
RoBERTa	0.702	0.705	0.722
FT-BERT	0.606	0.519	0.666
FT-RoBERTa	**0.506**	**0.480**	**0.485**

- Left/Right chain VS BERT/RoBERTa Induced Tree
- BERT/RoBERTa Induced Tree VS FT-BERT/RoBERTa Induced Tree

Analysis: Discussions

- **Aspects-sentiment Distance (AsD)**
  - the average distance between aspect and sentiment words.

$$AsD(S_i) = \frac{\sum_{w}^{w_i} \sum_{C'=S_i \cap C}^{C'_i} dist(C'_i, w_i)}{|w|\,|C'|}$$

$$AsD = \frac{\sum_{S}^{S_i} AsD(S_i)}{|S|}$$

- C sentiment words set
  - For Twitter —— pre-defined words set
  - For Rest14 and Laptop14 —— paired sentiment words to aspect （pAsD）

Tree Structure	Rest14	Laptop14	Twitter
Dep.	4.46 / 3.19	3.77 / 3.13	4.26
Left-chain	7.49 / 6.06	6.48 / 5.97	7.90
Right-chain	7.49 / 6.06	6.48 / 5.97	7.90
BERT	5.85 / 4.20	5.06 / 4.19	5.87
RoBERTa	5.05 / 3.61	4.49 / 3.67	5.39
FT-BERT	3.85 / 3.58	3.65 / 3.22	5.06
FT-RoBERTa	**3.56 / 2.92**	**3.35 / 2.88**	**3.55**

- BERT/RoBERTa Induced Tree VS FT-BERT/RoBERTa Induced Tree

More sentimentword-oriented!

Analysis: Surprise

FUDAN UNIVERSITY 復旦大學

Embedding	Model	Tree Structure	Rest14		Laptop14		Twitter	
			Acc.	F_1	*Acc.*	F_1	*Acc.*	F_1
Static Embedding	**BiLSTM** †	-	77.59	67.05	70.06	64.46	71.39	69.45
	LSTM+SynATT ♯	Dep.	80.45	71.26	72.57	69.13	-	-
	AdaRNN ♯	Dep.	-	-	-	-	66.30	65.90
	TD-GAT ♯	Dep.	80.35	76.13	74.13	72.01	72.68	71.15
BERT	**MLP**	-	85.35	78.38	78.36	74.16	75.92	74.41
	DGEDT ♯	Dep.	86.30	80.0	79.80	75.60	**77.90**	75.40
	RGAT ♯	Dep.	86.60	81.35	78.21	74.07	76.15	74.88
	RACL♯	-	-	**81.61**	-	73.91	-	81.61
RoBERTa	**MLP**	-	87.37	80.96	83.78	80.73	77.17	**76.20**
	RoBERTa-ASC ♯	Dep.	82.82	75.12	74.12	70.52	-	-
	LCFS-ASC-CDW ♯	Dep.	86.71	80.31	80.52	77.13	-	-
	ASGCN	Dep.	86.90	80.75	81.66	78.31	75.28	74.38
		FT-RoBERTa	86.87	80.59	83.33	80.32	76.10	75.07
	PWCN	Dep.	87.41	81.07	**84.16**	**81.18**	76.63	75.60
		FT-RoBERTa	87.35	80.85	84.01	81.08	77.02	75.52
	RGAT	Dep.	87.43	80.61	83.43	80.28	74.42	72.93
		FT-RoBERTa	**87.52**	81.29	83.33	79.95	75.81	74.91

Conclusion

What about the comparison between:

1. Tree induced from PTMs vs. Tree from dependency parser ?
    - Proportion of Neighboring Connections
    - Aspects-sentiment Distance (AsD)

2. Tree induced from PTMs vs. Tree from **task fine-tuned** PTMs ?
    - PTMs adapt the implicitly entailed tree structure during the finetuning
    - Tree from **task fine-tuned** PTMs is more sentiment-word-oriented even than the Tree given by parser

PTMs, YES！

Thanks !

Q & A

Paper link

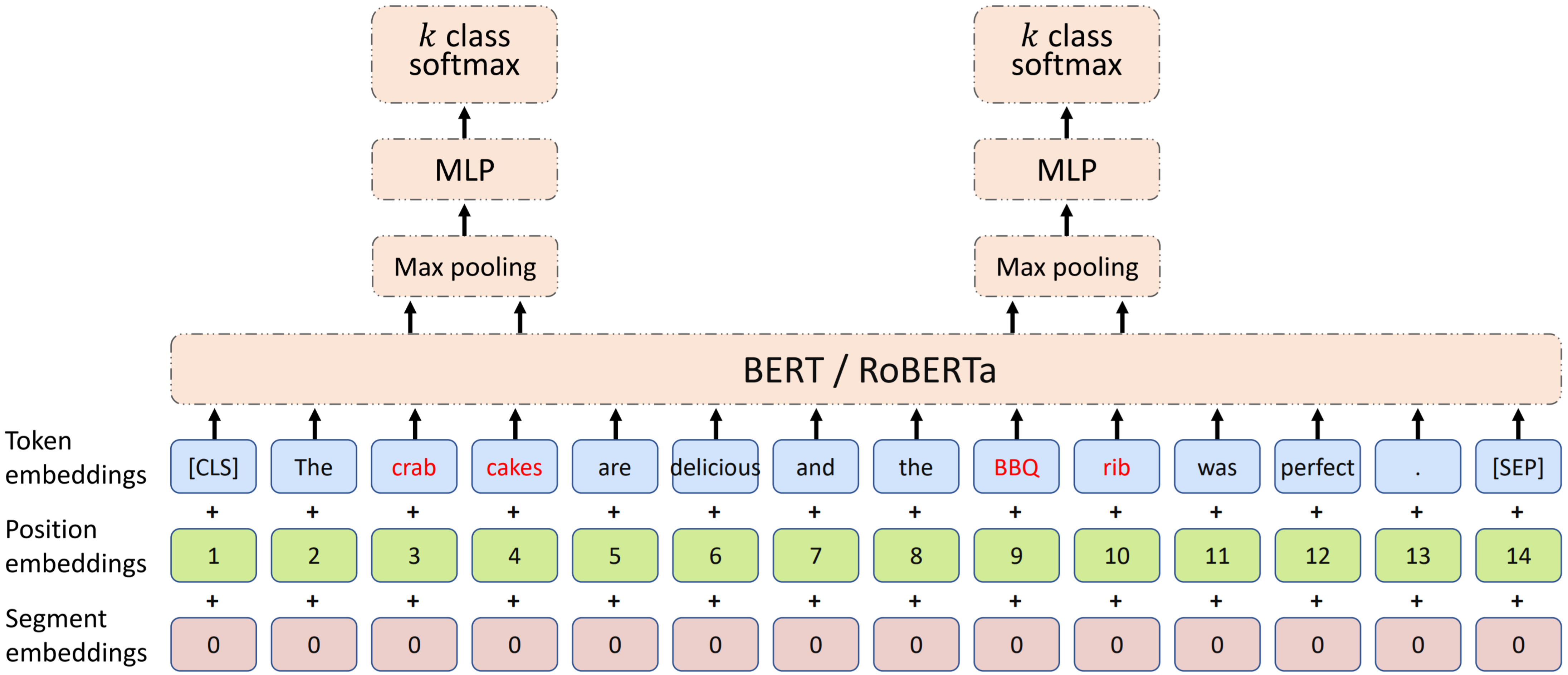
Github link

Paerwithcode

Explainaboard

jqdai19@fudan.edu.cn

NAACL 2021



Our Fine-tuning model