

IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality and Nugget Detection

¹Information Management, National Taipei University, New Taipei City, Taiwan

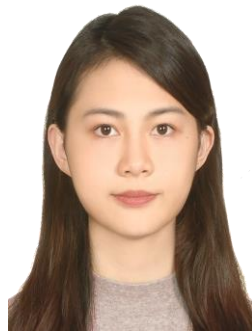
²Zeals Co., Ltd. Tokyo, Japan



Ting-Yun Hsiao¹



Yung-Wei Teng¹



Pei-Tz Chiu¹



Mike Tian-Jian Jiang²



Min-Yuh Day^{1,*}

myday@gm.ntpu.edu.tw

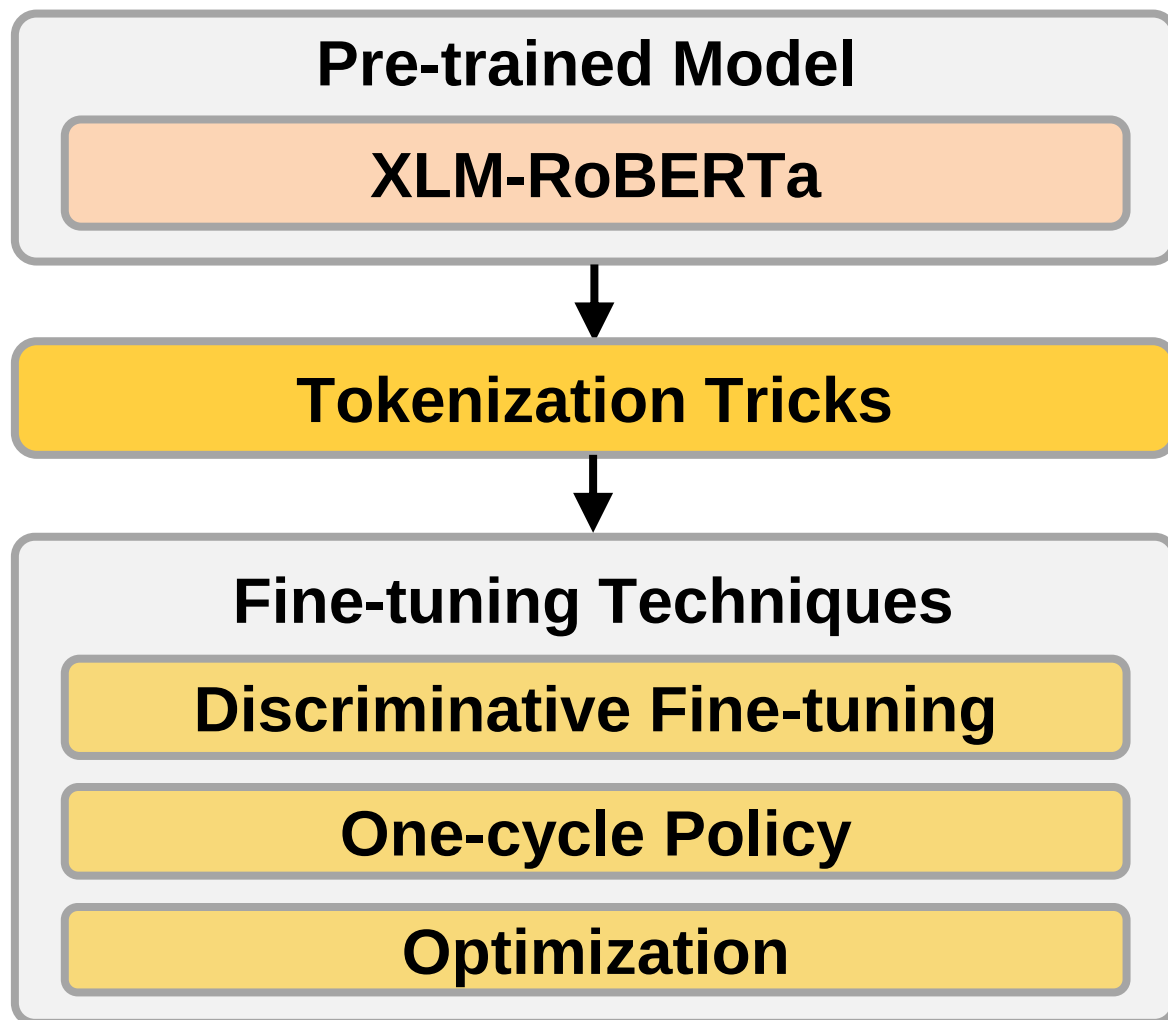
Outline

- **IMNTPU Research Architecture**
- **Proposed Approach**
 - **Tokenization Tricks**
 - **Fine-tuning Techniques**
- **Results**
- **Conclusion**

Highlights

- IMNTPU (**I**nformation **M**anagement at **N**ational **T**ai**P**ei **U**niversity)
Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue
Quality (DQ) and Nugget Detection (ND).
- IMNTPU submitted one run for Chinese DQ, English DQ and
English ND.

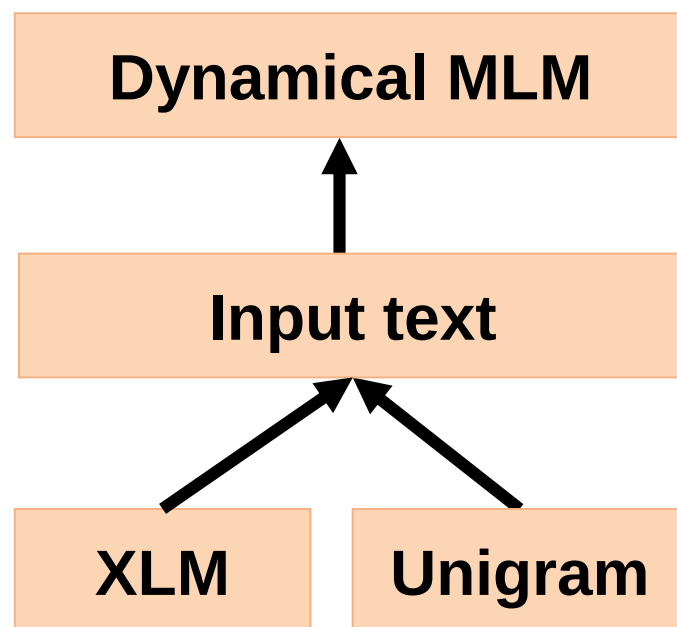
IMNTPU Research Architecture for NTCIR-16 DialEval-2



IMNTPU Research Architecture for NTCIR-16 DialEval-2

XLM-RoBERTa

Conneau et al. (2019)



- Retains subword token letter cases
- Integrates and improves approaches from **Cross-lingual Language Model (XLM)** pretraining schemes
- Tokenize with **unigram**-level sentence piece

Tokenization Tricks

BOS (beginning of sentence)

SEP (separator of sentences)

xxlen 3 <s> xxtrn 1 xxsdr customer Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service </s> </s> xxtrn 2 xxsdr helpdesk Hello, the problem of not displaying the lunar calendar in the view of the built-in calendar month will be updated in the later version. The external version of the calendar cannot display the dynamic icon at present. </s> </s> xxtrn 3 xxsdr customer I see. Thank you! </s>

EOS (end of sentence)

Tokenization Tricks

length

position

```
xxlen 3 <s> xxtrn 1 xxsdr customer Since there is no lunar  
calendar in the phone's calendar, I installed a new calendar  
application, but the date displayed is different. @Smartisan  
Customer Service </s> </s> xxtrn 2 xxsdr helpdesk Hello, the  
problem of not displaying the lunar calendar in the view of the  
built-in calendar month will be updated in the later version.  
The external version of the calendar cannot display the dynamic  
icon at present. </s> </s> xxtrn 3 xxsdr customer I see. Thank  
you! </s>
```

Tokenization Tricks

CNUGO

xxlen 3 <s> xxtrn 1 xxsdr customer Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service </s> </s> xxtrn 2 xxsdr helpdesk Hello, the problem of not displaying the lunar calendar in the view of the built-in calendar month will be updated in the later version. The external version of the calendar cannot display the dynamic icon at present. </s> </s> xxtrn 3 xxsdr customer I see. Thank you! </s>

Fine-tuning Techniques

Fine-tuning Techniques

Discriminative Fine-tuning

One-cycle Policy

Optimization

Fine-tuning Techniques

1

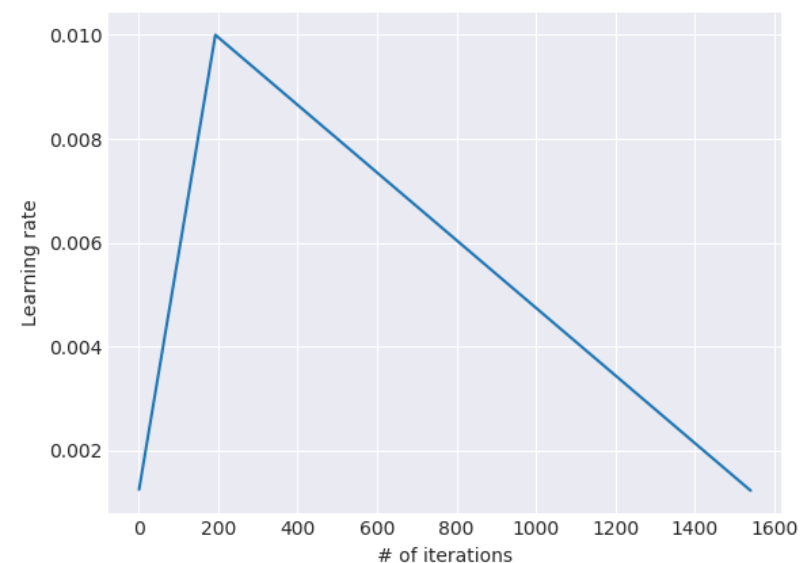
Discriminative Fine-tuning

- Different layers capture different types of information. They should be fine-tuned to different extents.
- The initial layers capture the most general form of information.
- General information of language are common and would require the least changes in their weights.
- The amount of fine-tuning required increases gradually as we move towards the last layer.

Fine-tuning Techniques

2 One-cycle Policy

- Slanted triangular learning rates
 - Intuition for adapting parameters to task-specific features.
 - The model should converge quickly to a suitable region and then refine its parameters.



Howard & Ruder (2018)

Results of IMNTPU at NTCIR-16 DialEval-2 Chinese Dialogue Quality

NTCIR-16 DialEval-2 Chinese Dialogue Quality (DQ) Test set						
	A-score		S-score		E-score	
Model	RSNOD	NMD	RSNOD	NMD	RSNOD	NMD
IMNTPU-run0	0.2479	0.1618	0.2032	0.1315	0.1860	0.1427
Baseline-run0	0.2301	0.1772	0.1998	0.1523	0.1854	0.1579
NTCIR-16 DialEval-2 Chinese Dialogue Quality (DQ) Development set						
	A-score		S-score		E-score	
Model	RSNOD	NMD	RSNOD	NMD	RSNOD	NMD
IMNTPU-run0	0.2262	0.1495	0.2076	0.1344	0.1694	0.1251

Results of IMNTPU at NTCIR-16 DialEval-2

English Dialogue Quality

NTCIR-16 DialEval-2 English Dialogue Quality (DQ) Test set						
	A-score		S-score		E-score	
Model	RSNOD	NMD	RSNOD	NMD	RSNOD	NMD
IMNTPU-run0	0.2535	0.1654	0.2020	0.1312	0.1826	0.1400
Baseline-run0	0.2321	0.1780	0.1986	0.1467	0.1745	0.1431
NTCIR-16 DialEval-2 English Dialogue Quality (DQ) Development set						
	A-score		S-score		E-score	
Model	RSNOD	NMD	RSNOD	NMD	RSNOD	NMD
IMNTPU-run0	0.2102	0.1397	0.1879	0.1216	0.1617	0.1184

Results of IMNTPU at NTCIR-16 DialEval-2 Chinese Nugget Detection

NTCIR-16 DialEval-2 Chinese Nugget Detection (ND) Test set

Model	JSD	RNSS
Baseline-run0	0.0585	0.1651

NTCIR-16 DialEval-2 Chinese Nugget Detection (ND) Development set

Model	JSD	RNSS
IMNTPU-run0	2.0670	1.3969

Results of IMNTPU at NTCIR-16 DialEval-2 English Nugget Detection

NTCIR-16 DialEval-2 English Nugget Detection (ND) Test set		
Model	JSD	RNSS
IMNTPU-run0	0.0601	0.1574
Baseline-run0	0.0625	0.1722
NTCIR-16 DialEval-2 English Nugget Detection (ND) Development set		
Model	JSD	RNSS
IMNTPU-run0	0.0752	0.1727

Conclusion

- **We employed XLM-RoBERTa to extract texts between conversations to learn key information and to cope with text categorization challenges in the DQ and ND subtasks.**

Contribution

- The important contribution of this study is that we proposed two critical elements, namely, **Tokenization** procedures and **Fine-Tuning Approaches**, to improve the DQ and ND subtasks in dialogue analysis.

IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality and Nugget Detection

¹Information Management, National Taipei University, New Taipei City, Taiwan

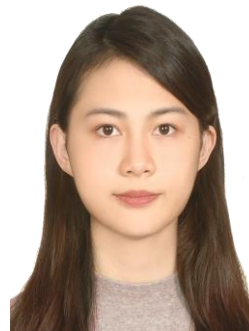
²Zeals Co., Ltd. Tokyo, Japan



Ting-Yun Hsiao¹



Yung-Wei Teng¹



Pei-Tz Chiu¹



Mike Tian-Jian Jiang²



Min-Yuh Day^{1,*}

myday@gm.ntpu.edu.tw