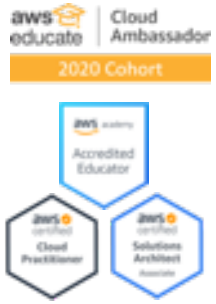
NTCIR Evaluation of Dialogue System and Financial Technology

Host: 吳怡瑾 教授 (Prof. Nancy I-Chin Wu)
Graduate Institute of Library & Information Studies, National Taiwan Normal University (NTNU)
Time: 13:30-15:30, September 16, 2022 (Friday)
Place: Graduate Institute of Library & Information Studies, NTNU



戴敏育 副教授

Min-Yuh Day, Ph.D, Associate Professor

國立臺北大學 資訊管理研究所

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2022-09-16

NTPU
國立臺北大學
National Taipei University

戴敏育 博士
(Min-Yuh Day, Ph.D.)

aws educate | Cloud Ambassador
2020 Cohort

aws academy
Accredited Educator

aws certified
Solutions Architect
Associate

aws certified
Cloud Practitioner

國立臺北大學 資訊管理研究所 副教授
中央研究院 資訊科學研究所 訪問學人
國立臺灣大學 資訊管理 博士

Publications Co-Chairs, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on Empirical Methods for Recognizing Inference in TExt (IEEE EM-RITE 2012-)

Publications Chair, The IEEE International Conference on Information Reuse and Integration for Data Science (IEEE IRI)

NTPU
國立臺北大學
National Taipei University

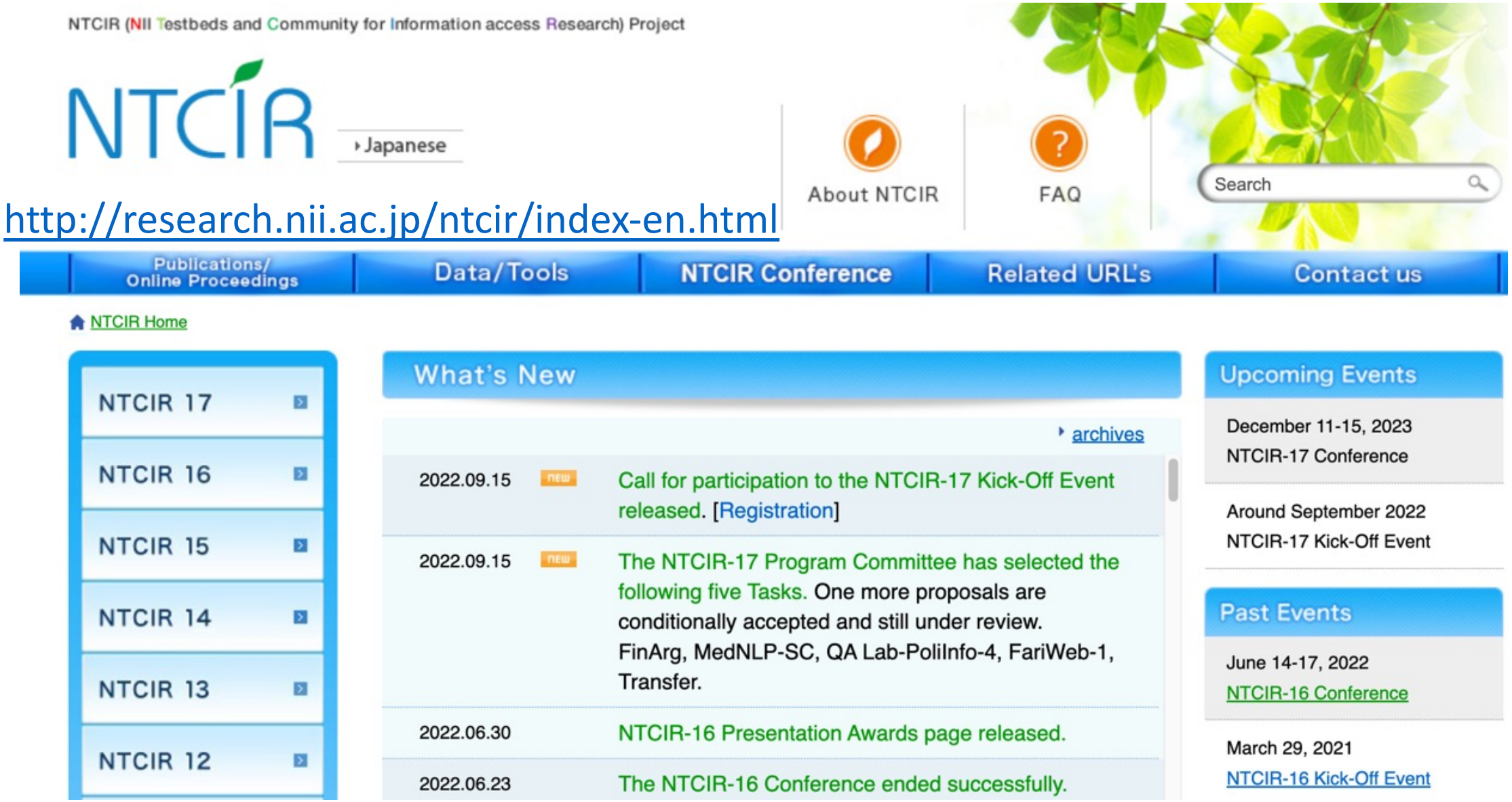
中央研究院 ACADEMIA SINICA 1928

國立臺灣大學

Outline

- **NTCIR Evaluation of Information Access Technologies**
  - **NTCIR 16 (2021-2022)**
  - **NTCIR 17 (2022-2023)**
- **Dialogue System**
- **Financial Technology**

NTCIR



http://research.nii.ac.jp/ntcir/index-en.html

Source: http://research.nii.ac.jp/ntcir/index-en.html

Why NTCIR Evaluation Forum?

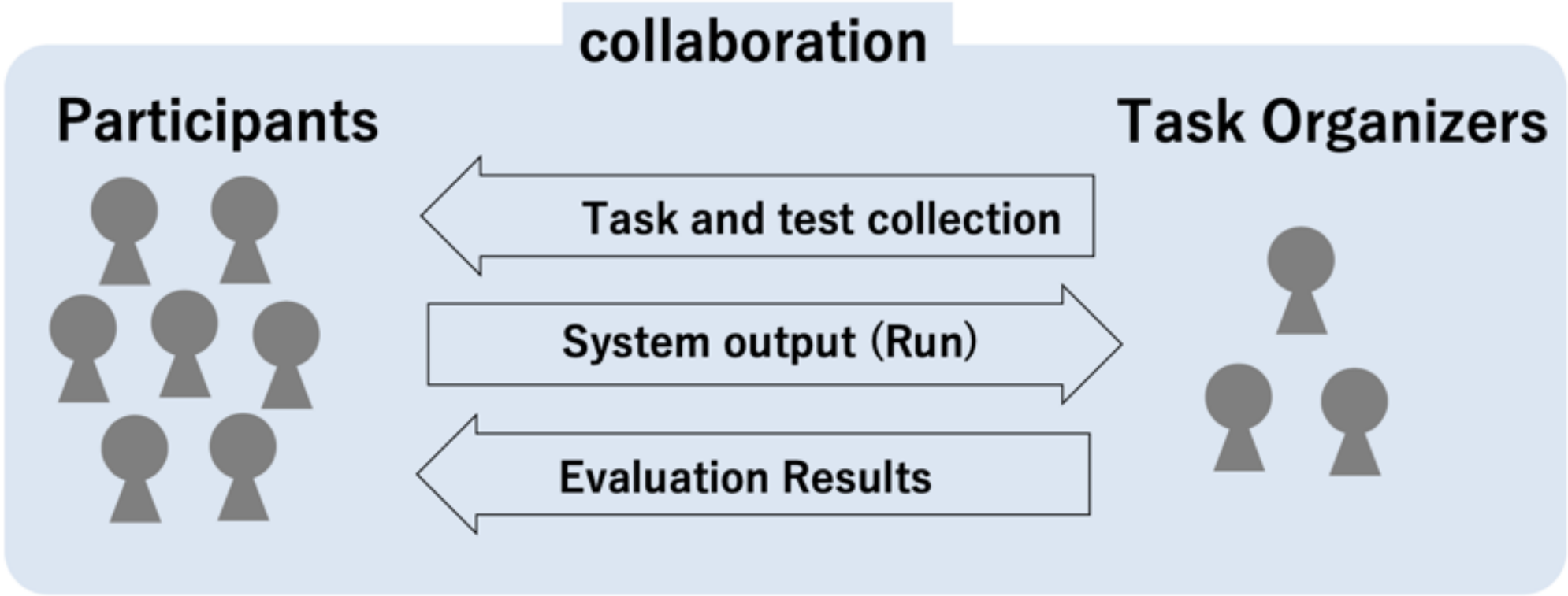
- **Can develop a large-scale test collections**
- **Can fairly compare the systems**
- **Can communicate with people who has the same research interest**

NII **T**estbeds and **C**ommunity for **I**nformation access **R**esearch

Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

NTCIR Evaluation Forum

- **Research groups gather together and tackle the same problems (problem definition, test collection, evaluation measure, etc.)**
- **The systems can be compared across participants**



Share results, exchange ideas

Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

History of NTCIR Tasks

Year	1999	2001	2002	2004	2005	2007	2008	2010	2011	2013	2014	2016	2017	2019	2020	2022
Task/NTCIR round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Total number	37	39	61	74	79	81	80	66	102	108	93	97	71	47	52	53
Automatic Term Recognition and Role Analysis (TMREC) (1)	9															
Ad hoc/Crosslingual IR (1) -> Chinese/English/Japanese IR (2) -> CLIR (3-6)	28	30	20	26	25	22										
Text Summarization Challenge (TSC) (2-4)		9	8	9												
Web Retrieval (WEB) (3-5)			7	11	7											
Question Answering Challenge (QAC) (3-6)			16	18	7	8										
Patent Retrieval [and Classification] (PATENT) (3-6)			10	10	13	12										
Multimodal Summarization for Trend Information (MUST) (5-7)					13	15	13									
Crosslingual Question Answering (CLQA) (5, 6) -> Advanced Crosslingual Information Access (ACLIA) (7, 8)					14	12	19	14								
Opinion (6) -> Multilingual Opinion Analysis (MOAT) (7, 8)						12	21	16								
Patent Mining (PAT-MN) (7, 8)							12	11								
Community Question Answering (CQA) (8)								4								
Geotemporal IR (GeoTime) (8, 9)								13	12							
Interactive Visual Exploration (Vis-Ex) (9)									4							
Patent Translation (PAT-MT)(7, 8) -> Patent Machine Translation (PatentMT)(9, 10)							15	8	21	21						
Crosslingual Link Discovery (Crosslink) (9, 10)									11	10						
INTENT(9, 10) -> Search Intent and Task Mining (IMine) (11, 12)									16	11	12	9				
One Click Access (1CLICK)(9, 10) -> Mobile Information Access (MobileClick) (11, 12)									4	8	4	11				
Recognizing Inference in Text (RITE)(9,10) -> Recognizing Inference in Text and Validation (RITE-VAL)(11)									24	28	23					
IR for Spoken Documents (SpokenDoc) (9, 10) -> Spoken Query and Spoken Document Retrieval (SpokenQuery&Doc) (11, 12)									10	12	11	7				
Mathematical Information Access (Math) (10, 11) -> MathIR (12)										6	8	6				
Medical Natural Language Processing (MedNLP) (10, 11) -> MedNLPDoc (12) -> MedWeb (13) -> Real-MedNLP(16)										12	12	8		9		10
QA Lab for Entrance Exam (QALab) (11, 12, 13) -> QA Lab for Political Information (QALab-PoliInfo) (14, 15, 16)											11	12	11	13	14	12
Temporal Information Access (Temporalia) (11, 12)											8	14				
Cooking Recipe Search (RecipeSearch) (11)											4					
Personal Lifelog Organisation & Retrieval (Lifelog) (12, 13, 14, 16)												8	4	6		3
Short Text Conversation (STC) (12, 13, 14)												22	27	13		
Open Live Test for Question Retrieval (OpenLiveQ) (13, 14)													7	4		
Actionable Knowledge Graph (AKG) (13)													3			
Emotion Cause Analysis (ECA) (13)													3			
Neurally Augmented Image Labelling Strategies (NAILS) (13)													2			
We Want Web (WWW) (13, 14) -> We Want Web with CENTER (WWW) (15, 16)													5	4	8	3
Fine-Grained Numeral Understanding in Financial Tweet (FinNum) (14,15,16)														6	7	7
CLEF/NTCIR/TREC REproducibility (CENTRE) (14)														1		
Dialogue Evaluation (DialEval) (15, 16)															7	4
SHINRA 2020 Multi-lingual (SHINRA 2020-ML) (15)															7	
Data Search (Data Search) (15, 16)															5	6
Micro Activity Retrieval Task (MART) (15)															5	
Session Search (SS) (16)																3
Reading Comprehension for Information Retrieval (RCIR) (16)																3
Unbiased Learning to Ranking Evaluation Task (ULTRE) (16)																2

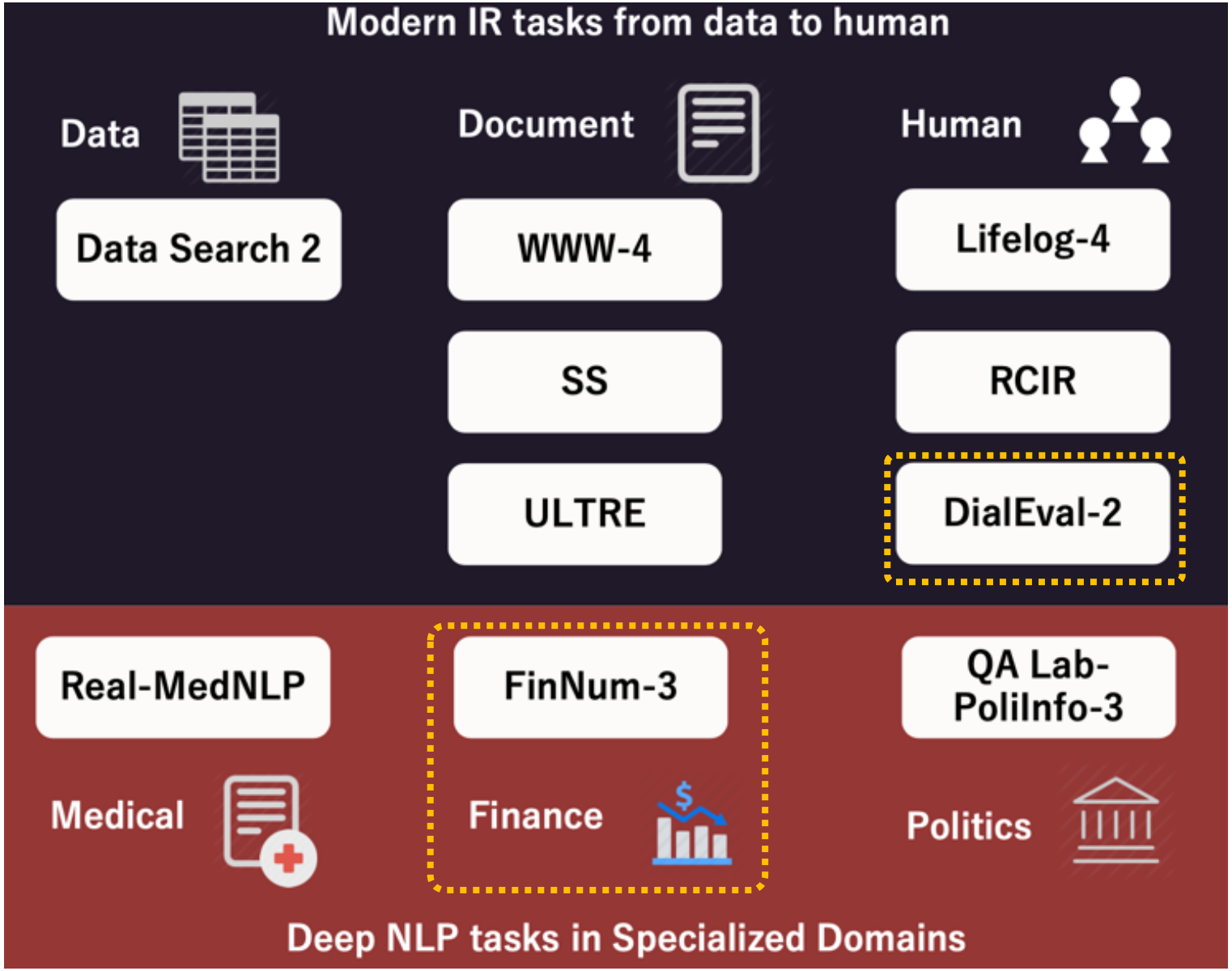
Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

History of NTCIR-16 Tasks

Year	2013	2014	2016	2017	2019	2020	2022
Task/NTCIR round	10	11	12	13	14	15	16
Medical Natural Language Processing (MedNLP) (10, 11) -> MedNLPDoc (12) -> MedWeb (13) –> **Real-MedNLP**(16)	12	12	8	9			10
QA Lab for Entrance Exam (QALab) (11, 12, 13) -> QA Lab for Political Information (**QALab-PoliInfo**) (14, 15, 16)		11	12	11	13	14	12
Personal Lifelog Organisation & Retrieval (**Lifelog**) (12, 13, 14, 16)			8	4	6		3
We Want Web (WWW) (13, 14) -> We Want Web with CENTER (**WWW**) (15, 16)				5	4	8	3
Fine-Grained Numeral Understanding in Financial Tweet (**FinNum**) (14,15,16)					6	7	7
Dialogue Evaluation (**DialEval**) (15, 16)						7	4
Data Search (**Data Search**) (15, 16)						5	6
Session Search (**SS**) (16)							3
Reading Comprehension for Information Retrieval (**RCIR**) (16)							3
Unbiased Learning to Ranking Evaluation Task (**ULTRE**) (16)							2

Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

Information Retrieval + NLP



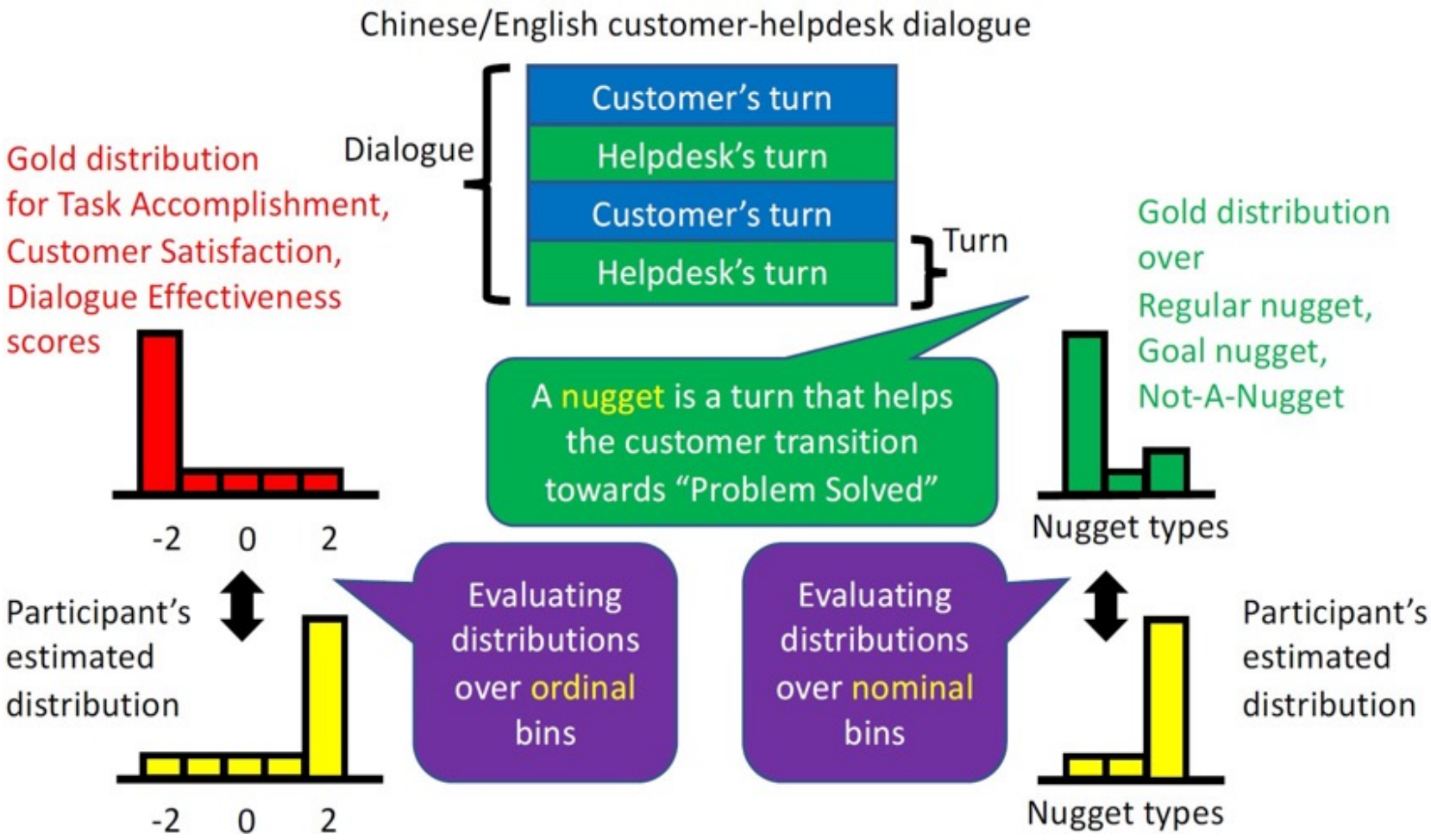
Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

DialEval-2 (Dialogue Evaluation)

Evaluation of the quality of a customer-helpdesk dialogue

Dialogue Quality Subtask

Nugget Detection Subtask



Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

FinNum-3

(Investor's and Manager's Fine-grained Claim Detection)

Fine-grained understanding of claims in financial documents

Investors always make a claim with an estimation

We estimate that the sales growth rate may exceed 80%

- Detect the given numeral is in-claim or out-of-claim
- Classify the category of the numeral
- Documents: Investors reports, earnings conference calls

Source: Takehiro Yamamoto and Zhicheng Dou (2022), Overview of NTCIR-16, NTCIR-16 Proceedings

The 17th NTCIR (2022 - 2023)

Evaluation of Information Access Technologies

July 2022 - December 2023

- **CORE TASKS**
  - **Fine-grained Argument Understanding in Financial Analysis (FinArg)**
  - **Medial Natural Language Processing for Social media and Clinical texts (MedNLP-SC)**
  - **QA Lab for Political Information 4 (QA Lab-PoliInfo-4)**
- **PILOT TASKS**
  - **The First Fair Web Task (FairWeb-1)**
  - **Resource Transfer Based Dense Retrieval (Transfer)**

Source: https://research.nii.ac.jp/ntcir/ntcir-17/index.html

The 17th NTCIR (2022 - 2023)

Evaluation of Information Access Technologies

July 2022 - December 2023

- Sep 2022 NTCIR-17 kickoff
- Jan 2023 Dataset Release*
- Jan-Jun, 2023 Dry Run*
- Mar-Jul, 2023 Formal Run*
- Aug 1, 2023 Evaluation result return
- Aug 1, 2023 Task overview paper release (draft)
- Sep 1, 2023 Submission due of participant papers (draft)
- Nov 1, 2023 Camera-ready participant paper due
- Dec 2023 NTCIR-17 Conference in NII, Tokyo

Source: https://research.nii.ac.jp/ntcir/ntcir-17/index.html

Tamkang University
淡江大學

2011

Tamkang University

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-9 RITE

Department of Information Management
Tamkang University, Taiwan

Min-Yuh Day **Chun Tu**

myday@mail.tku.edu.tw

NTCIR-9 Workshop, December 6-9, 2011, Tokyo, Japan

Tamkang University
淡江大學

2013

Tamkang University

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-10 RITE-2

Department of Information Management
Tamkang University, Taiwan

Min-Yuh Day

Chun Tu

Hou-Cheng Vong

Shih-Wei Wu

Shih-Jhen Huang

myday@mail.tku.edu.tw

NTCIR-10 Conference, June 18-21, 2013, Tokyo, Japan

IMTKU Textual Entailment System for Recognizing Inference in Text at NTCIR-11 RITE-VAL 2014

Tamkang University

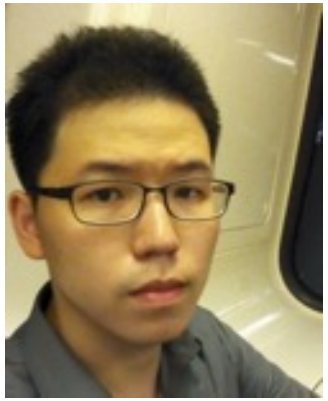
淡江大學

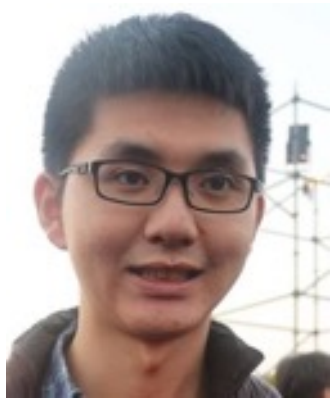

Min-Yuh Day

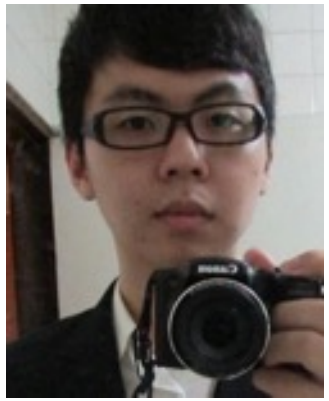

Ya-Jung Wang


Che-Wei Hsu


En-Chun Tu


Huai-Wen Hsu


Yu-An Lin


Shang-Yu Wu


Yu-Hsuan Tai


Cheng-Chia Tsai

NTCIR

Tamkang University
淡江大學

2016

IMTKU Question Answering System for World History Exams at NTCIR-12 QA Lab2

Department of Information Management
Tamkang University, Taiwan

Sagacity Technology


Min-Yuh Day


Cheng-Chia Tsai


Wei-Chun Chung


Hsiu-Yuan Chang


Tzu-Jui Sun


Yuan-Jie Tsai


Jin-Kun Lin


Cheng-Hung Lee


Yu-Ming Guo


Yue-Da Lin


Wei-Ming Chen


Yun-Da Tsai


Cheng-Jhih Han


Yi-Jing Lin


Yi-Heng Chiang


Ching-Yuan Chien

myday@mail.tku.edu.tw

NTCIR-12 Conference, June 7-10, 2016, Tokyo, Japan

NTCIR

2017

Tamkang University
淡江大學

IMTKU Question Answering System for World History Exams at NTCIR-13 QALab-3

Department of Information Management
Tamkang University, Taiwan

Min-Yuh Day

Chao-Yu Chen

Wanchu Huang

Shi-Ya Zheng

I-Hsuan Huang

Tz-Rung Chen

Min-Chun Kuo

Yue-Da Lin

Yi-Jing Lin

myday@mail.tku.edu.tw

NTCIR-13 Conference, December 5-8, 2017, Tokyo, Japan

NTCIR

2019

Tamkang University
淡江大學

IMTKU Emotional Dialogue System for Short Text Conversation at NTCIR-14 STC-3 (CECG) Task

Department of Information Management
Tamkang University, Taiwan


Min-Yuh Day


Chi-Sheng Hung


Yi-Jun Xie


Jhih-Yi Chen


Yu-Ling Kuo


Jian-Ting Lin

myday@mail.tku.edu.tw

NTCIR-14 Conference, June 10-13, 2019, Tokyo, Japan

NTCIR

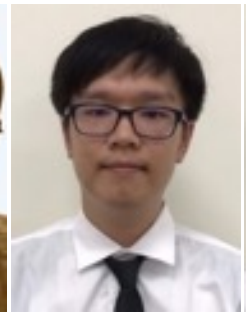
2020

Tamkang University 淡江大學

IMTKU Multi-Turn Dialogue System Evaluation at the NTCIR-15 DialEval-1 Dialogue Quality and Nugget Detection

[1] Zeals Co., Ltd. Tokyo, Japan
[2] Information Management, Tamkang University, Taiwan
[3] Information Management, National Taipei University, Taiwan


Mike Tian-Jian Jiang[1]

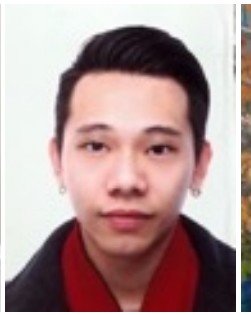

Zhao-Xian Gu[2]


Cheng-Jhe Chiang[2]


Yueh-Chia Wu[2]


Yu-Chen Huang[2]


Cheng-Han Chiu[2]


Sheng-Ru Shaw[2]


Min-Yuh Day[3]

NTCIR

2022

AACSB ACCREDITED

NTPU 國立臺北大學 National Taipei University

zeals

IMNTPU at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial Numclaim Classification

[1] Information Management, National Taipei University, New Taipei City, Taiwan
[2] Zeals Co., Ltd. Tokyo, Japan





Yung-Wei Teng [1] Pei-Tz Chiu [1] Ting-Yun Hsiao [1] Mike Tian-Jian Jiang [2] Min-Yuh Day [1,*]

myday@gm.ntpu.edu.tw

2022

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

[1] Information Management, National Taipei University, New Taipei City, Taiwan
[2] Zeals Co., Ltd. Tokyo, Japan





Ting-Yun Hsiao [1] Yung-Wei Teng [1] Pei-Tz Chiu [1] Mike Tian-Jian Jiang [2] Min-Yuh Day [1,*]

myday@gm.ntpu.edu.tw

NTCIR-16 Best Oral Presentation Award Recipients

2022

NTCIR-16 Best Oral Presentation Award Recipients

Task	Recipients
Data Search 2	*Lya Hulliyyatus Suadaa, Lutfi Rahmatuti Maghfiroh, Muhammad Luqman and Isfan Nur Fauzi.* STIS at the NTCIR-16 Data Search 2 Task: Ad-hoc Data Retrieval Ranking with Pretrained Representative Words Prediction
DialEval-2	*Fan Li and Tetsuya Sakai.* RSLDE at the NTCIR-16 DialEval-2 Task
FinNum-3	*Yung-Wei Teng, Pei-Tz Chiu, Ting-Yun Hsiao, Mike Tian-Jian Jiang and Min-Yuh Day.* IMNTPU at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial Numclaim Classification
Lifelog-4	*Ly Duyen Tran, Manh Duy Nguyen, Nhu Nguyen Thao, Cathal Gurrin, Minh-Triet Tran, Binh Nguyen, Liting Zhou.* HCMUS & DCU at NTCIR16-Lifelog4
QA Lab-PoliInfo-3	*Yasuhiro Ogawa, Yugo Kato and Katsuhiko Toyama.* NUKL's QA System at the NTCIR-16 QA Lab-PoliInfo-3

Source: https://research.nii.ac.jp/ntcir/ntcir-16/awards.html

NTCIR-16 Best Poster Presentation Award Recipients

2022

NTCIR-16 Best Poster Presentation Award Recipients

Task	Recipients
Data Search 2	*Lya Hulliyyatus Suadaa, Lutfi Rahmatuti Maghfiroh, Muhammad Luqman, Isfan Nur Fauzi.* STIS at the NTCIR-16 Data Search 2 Task: Ad-hoc Data Retrieval Ranking with Pretrained Representative Words Prediction
DialEval-2	*Fan Li, Tetsuya Sakai.* RSLDE at the NTCIR-16 DialEval-2 Task
FinNum-3	*Yung-Wei Teng, Pei-Tz Chiu, Ting-Yun Hsiao, Mike Tian-Jian Jiang, Min-Yuh Day.* IMNTPU at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial Numclaim Classification
Lifelog-4	*Naushad Alam, Ahmed Alateeq, Yvette Graham, Mark Roantree, Cathal Gurrin.* DCU at the NTCIR16 Lifelog-4 Task
QA Lab-PoliInfo-3	*Kazuma Kadowaki, Shunsuke Onuma.* JRIRD at the NTCIR-16 QA Lab-PoliInfo-3 Budget Argument Mining

Source: https://research.nii.ac.jp/ntcir/ntcir-16/awards.html

yahoo! 新聞 台灣新生報

臺北大學 國際發光

【記者王志誠、周貞伶／新北報導】

2022年7月9日 週六 下午8:33

由臺北大學資管所戴敏育副教授領軍的「IMNTPU」跨國團隊，在第十六屆NTCIR國際資訊存取技術評估研討會上榮獲多項大獎。其中在投資者與管理者的細粒度聲明檢測的中文分析報告分項與對話系統評測（DialEval-2）的英文金塊偵測分項EnglishNuggetDetection（ND）子任務，兩項子任務皆拿下第一名的優秀成績。

國立臺北大學資管所在戴敏育副教授帶領IMNTPU跨國團隊，其成員包括資管所碩士班研究生鄧詠薇、邱沛慈與蕭婷云，以及與日本東京Zeals公司AI自然語言科學家姜天戩共同合作，參與2022 NTCIR-16研討會榮獲許多獎項，為臺北大學資管所在NTCIR研討會上，建立良好的國際聲譽。

臺北大學資管所IMNTPU團隊在投資者與管理者的細粒度聲明檢測任務（FinNUM3）中，最終在七支隊伍中脫穎而出，除了在任務中取的平均效能為所有隊伍中最佳榮獲第一名，還囊括多項大獎，包含「口頭報告獎」與「海報展示獎」。不僅在FinNUM3口頭報告中以優秀的國際簡報與問答獲得主辦單位的高度賞識與重視；在海報展示期間，以精美海報展示與生動活潑的解說，吸引大批與會人員前來駐足交流與提問，獲得超高人氣。

戴敏育表示，希望藉由此次在國際研討會的優良成果，鼓勵學生積極參與相關國際競賽，讓學生能與國際接軌。IMNTPU隊長鄧詠薇認為，整個NTCIR-16比賽從拿到資料集、模型建置到最後預測結果，花了近半年的時間，突破重重關卡，到最後甚至能順利的在研討會中發表，這一過程受益良多。從主辦方提供的專業財務分析報告資料集，進一步針對聲明內容作細粒度分析，判斷聲明內容中的數字是否為其中重要資訊，以利相關利益者更能了解數字對於專業財務報告的重要性。

副隊長邱沛慈更談到，能夠在NTCIR-16 FinNUM中文財務分析報告中獲得第一名的績效，除了團隊共同努力外，也非常感謝戴老師與姜天戩博士在過程中給予很多建議與幫助。「經過約半年的努力，IMNTPU團隊在NTCIR-16 Dial-Eval-2對話系統評測任務英文金塊偵測分項能獲得第一名，真的很高興。」副隊長蕭婷云認為，從參與國際資訊競賽的過程中，可學習到許多寶貴經驗。

校方表示，在此次競賽中，透過與來自世界各國的高手較量，展現北大資管所的研究成果，不僅能開闊國際視野，也同時能讓世界各國看到臺灣隊伍的實力。

Source: https://tw.news.yahoo.com/%E8%87%BA%E5%8C%97%E5%A4%A7%E5%AD%B8-%E5%9C%8B%E9%9A%9B%E7%99%BC%E5%85%89-123346251.html

NTCIR

2022

AACSB ACCREDITED

NTPU 國立臺北大學 National Taipei University

zeals

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

[1] Information Management, National Taipei University, New Taipei City, Taiwan
[2] Zeals Co., Ltd. Tokyo, Japan





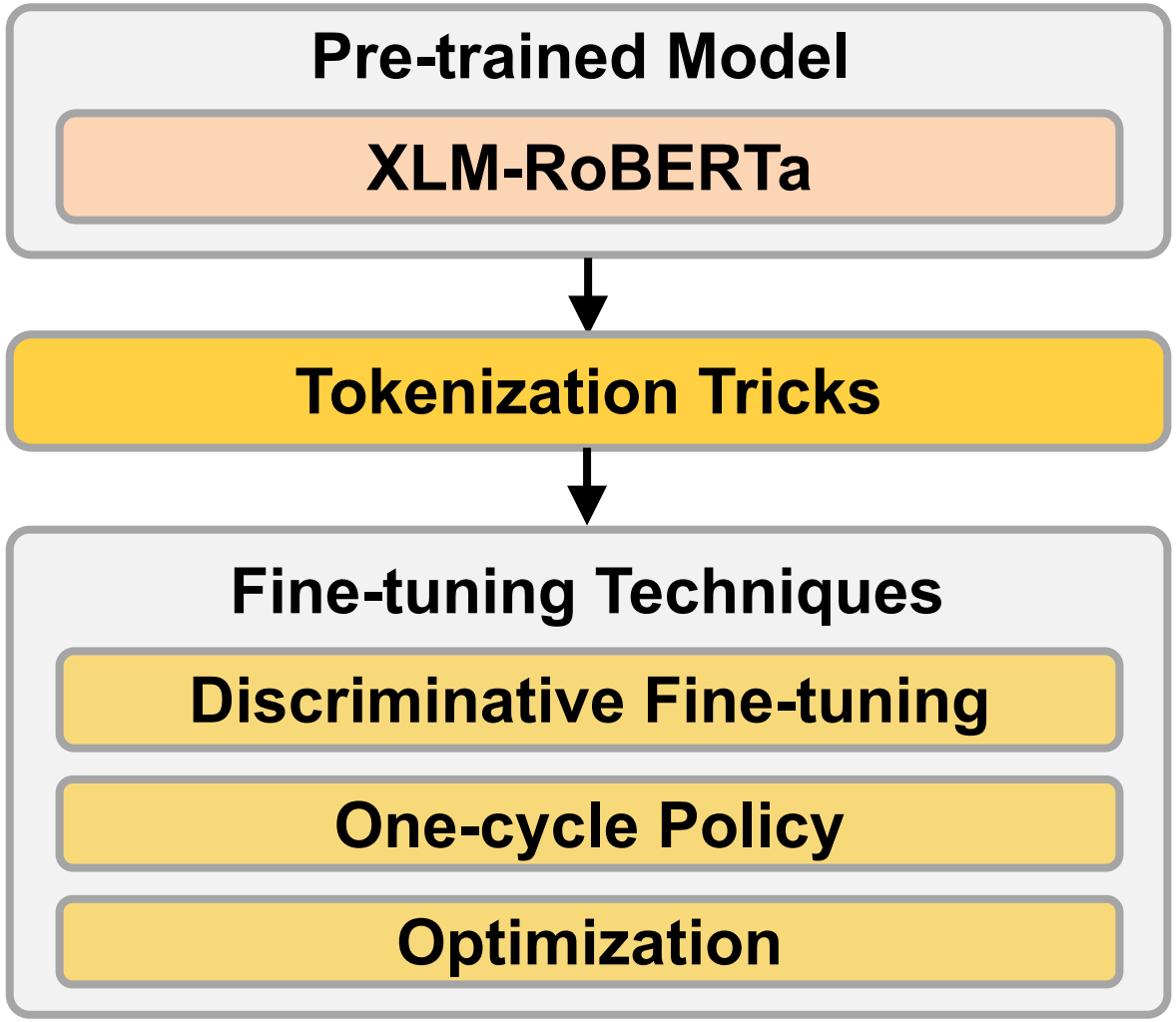
Ting-Yun Hsiao [1] Yung-Wei Teng [1] Pei-Tz Chiu [1] Mike Tian-Jian Jiang [2] Min-Yuh Day [1,*]

myday@gm.ntpu.edu.tw

Highlights

- **IMNTPU (Information Management at National TaiPei University) 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.**

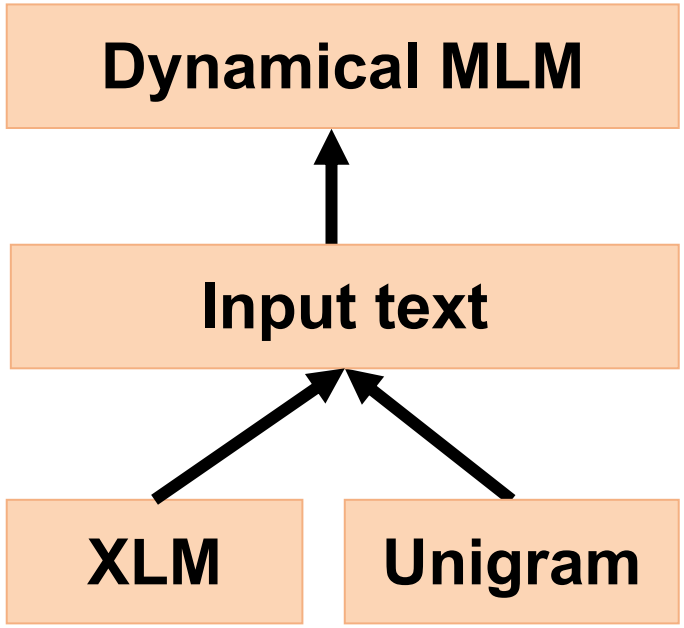
IMNTPU Research Architecture for NTCIR-16 DialEval-2



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**

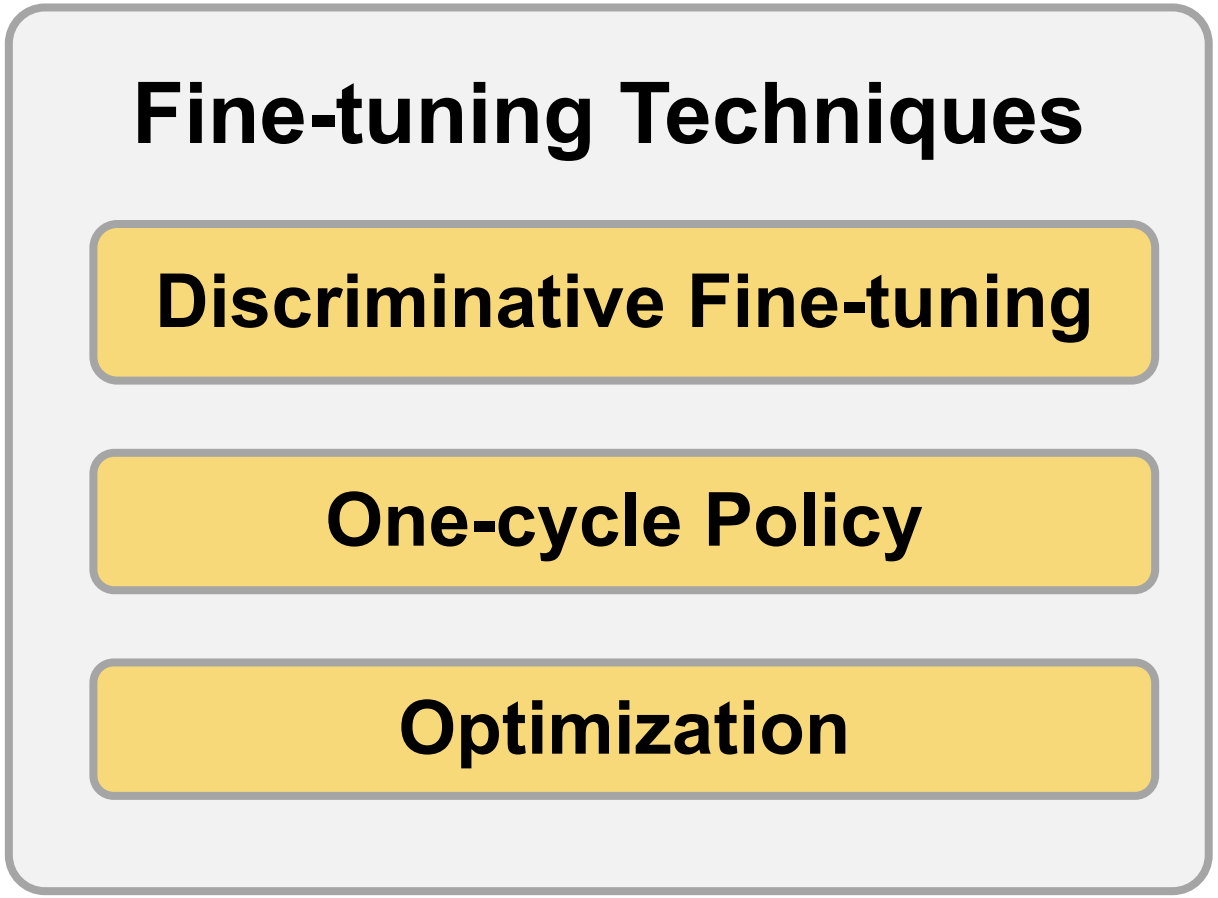
xxlen 3 <s> xxtrn 1 xxsdr customer Since there is no lunar calendar in the phone's calendar, I installed a new calendar application, but the date displayed is different. @Smartisan Customer Service </s> </s> xxtrn 2 xxsdr helpdesk Hello, the problem of not displaying the lunar calendar in the view of 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

CNUG0

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

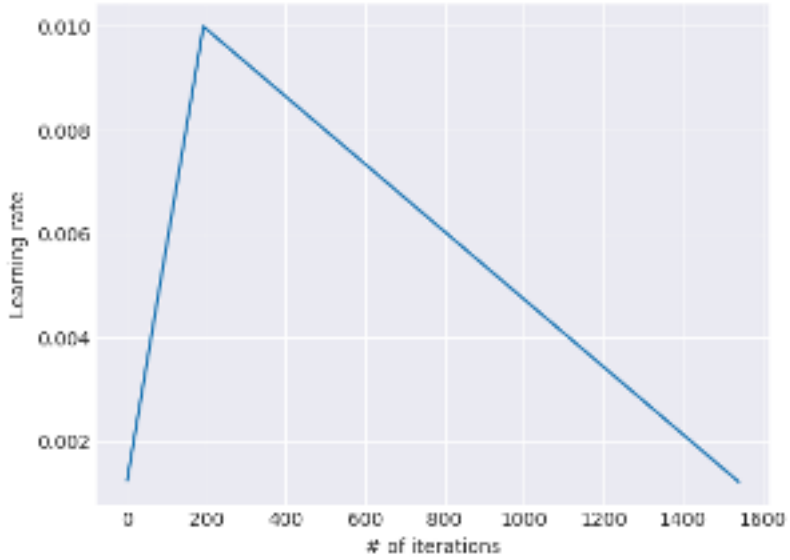
1 Discriminative Fine-tuning

- 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 paramaters.



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.**

NTCIR

2022

AACSB ACCREDITED

NTPU 國立臺北大學 National Taipei University

zeals

IMNTPU at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial Numclaim Classification

[1] Information Management, National Taipei University, New Taipei City, Taiwan
[2] Zeals Co., Ltd. Tokyo, Japan





Yung-Wei Teng [1] Pei-Tz Chiu [1] Ting-Yun Hsiao [1] Mike Tian-Jian Jiang [2] Min-Yuh Day [1,*]

myday@gm.ntpu.edu.tw

Highlights

- **IMNTPU**

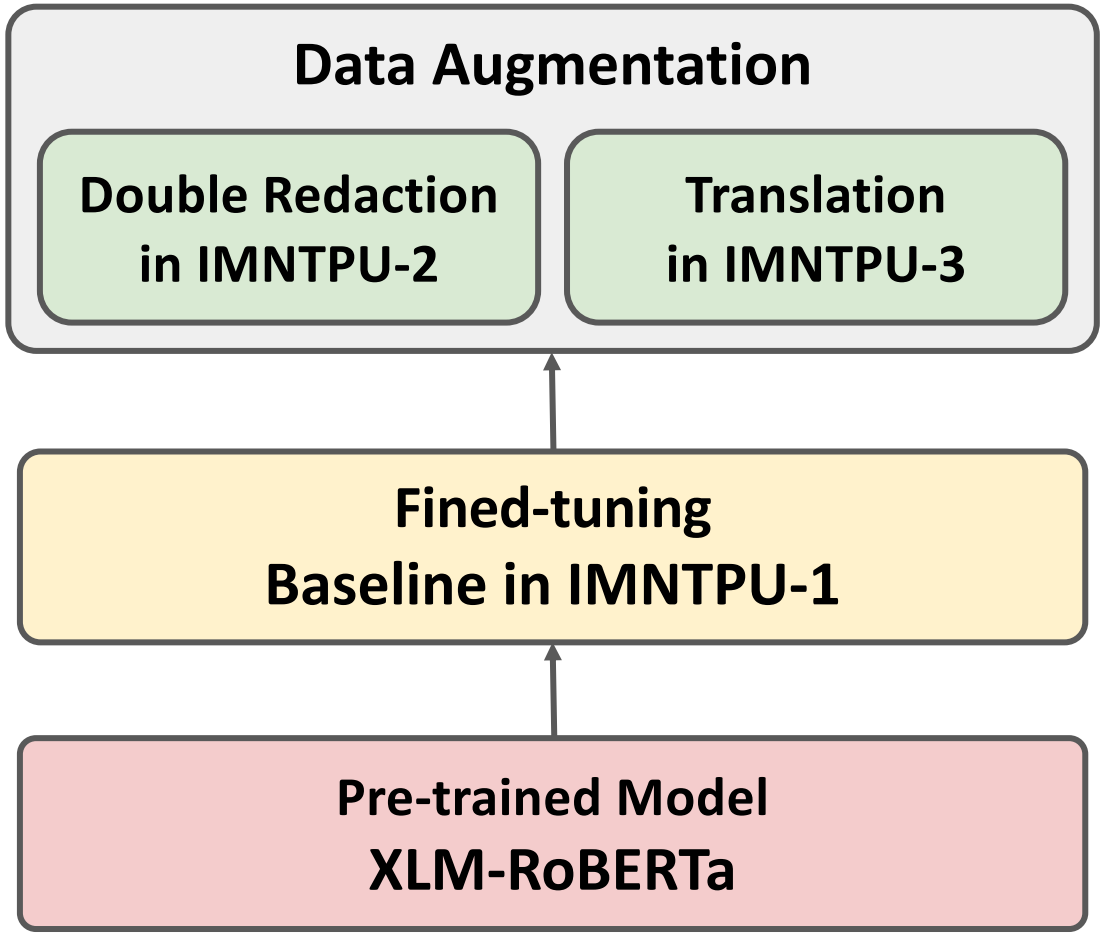
  (Information Management at National Taipei University)

  at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial

  Numclaim Classification

- IMNTPU Submitted **Three runs** for NTCIR-16 FinNUM3
  - IMNTPU1- XLMRoBERTa Baseline Model
  - IMNTPU2- Double Redaction
  - IMNTPU3- Translation

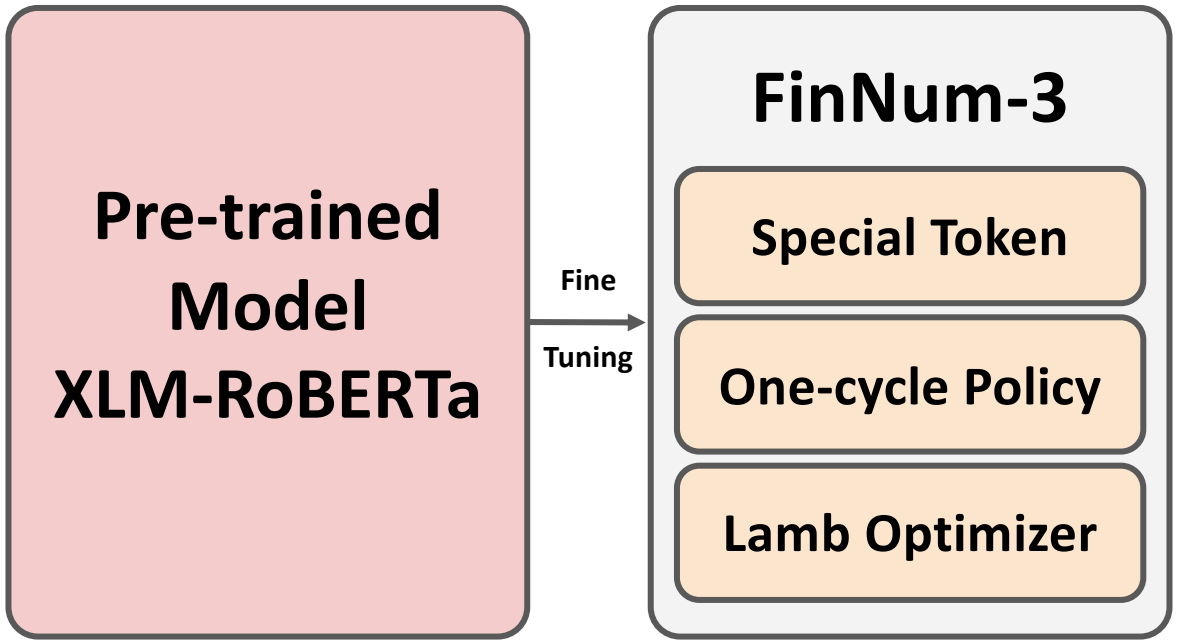
IMNTPU Research Architecture for NTCIR-16 FinNum-3



Proposed Method

- IMNTPU1 - We adopted **XLM-RoBERTa Model** without data augmentation as our baseline model.
- IMNTPU2 - We adopt **Double Redaction approach** for data augmentation in XLM-RoBERTa Model.
- IMNTPU3 - We adopt **Translation approach** for data augmentation in XLM-RoBERTa Model.

Fine-tuning of XLM-RoBERTa for IMNTPU at FinNum-3



- **Combine** :

  cross-lingual language model (XLM)
- **Tokenizer** : add Special Token
- **Optimizer** : Lamb Optimizer
- **Learning Rate** : One-Cycle Policy

Input:

Good day and welcome to the Apple Inc. Third Quarter Fiscal Year 2018 Earnings Conference Call. Today's call is being recorded.

XLM-RoBERTa Tokenizer

Output:

<s> Good day and welcome to the Apple Inc. Third Quarter Fiscal Year xxnum 2018 Earnings Conference Call. Today's call is being recorded. </s>

IMNTPU2- Double Redaction

Algorithm 1 An algorithm of double redaction

```
Shuffle the tokens in sentence
Delete the duplicated tokens in sentence
Copy the remaining tokens as β
SET the δ and γ
for specific token in β do
    if γ less than δ then
        Replace original token with <usk> token
    else
        Cover original token as <mask> token
    end if
end for
while True do
    Model predict the original token of <usk> and <mask>
end while
```

Double Redaction- English

Input:

Good day and welcome to the Apple Inc. Third Quarter Fiscal Year 2018 Earnings Conference Call. Today's call is being recorded.

Double Redaction for Data Augmentation

Output:

<s> <mask> day and <mask> to the Apple <mask> <mask> Quarter Fiscal Year xxnum 2018 Earnings Conference Call. Today's call is <mask> recorded. </s>

Double Redaction- Chinese

Input:

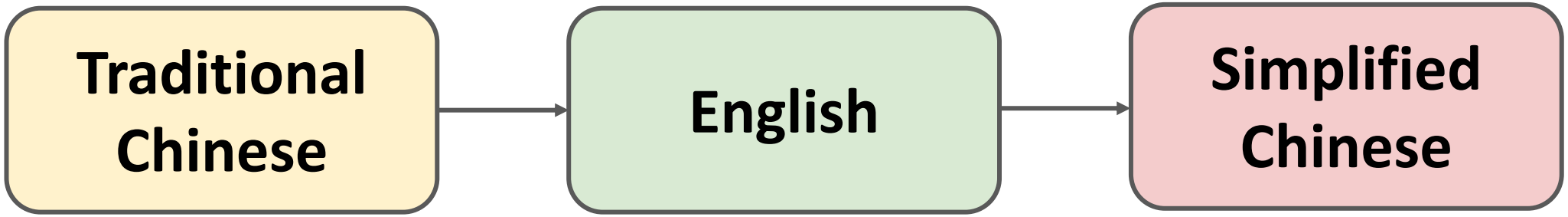
巨大為全球最大自行車製造商，擁有捷安特、Liv、Momentum三個自有品牌，營收比重 70%；代工業務佔 30%，最大客戶為 TREK。主要競爭優勢在生產規模龐大，創造了成本優勢，也使其生產工藝不斷精進。品牌經營則有多面向且細膩的操作經驗，2000年和品牌顧問公司 Interbrand 合作，希望用新品牌精神：啟動探索的熱情(InspiringAdventure)連結消費者，開始各項運動行銷操作。

Double Redaction for Data Augmentation

Output:

<mask> 巨大為全球最大自行車製造商，擁有捷安特、Liv、Momentum 三個自有品牌，營收比重 70%；代工業務佔 xxnum 30%，最大客戶為 TREK。主要競爭優勢在生產規模龐大，創造了成本優勢，也使其生產工藝不斷精進。品牌經營則有多面向且細膩的操作經驗，2000 年和品牌顧問公司 Interbrand 合作，希望用新品牌精神：啟動探索的熱情(InspiringAdventure)連結消費者，開始各項運動行銷操作。 </s>

IMNTPU3- Translation



Traditional Chinese	English	Simplified Chinese
“ 稅後純益 9.81 億元，YoY+36.36-%，稅後 EPS2.62 元，優於預期。”	“The tax proceeds were $981 million, YoY+36.36 percent and EPS 2.62 percent, higher than expected.”	“ 税后净利润为 9.81 亿美元，YoY+36.36%，扣除 ESP 2.62 税后利润比预期的要高。”

Performance- Chinese

Run	Dev Set F1-Score(%)	Test Set F1-Score(%)
IMNTPU-1 (Baseline)	90.51	**93.18**
IMNTPU-2 (Double Redaction)	88.65	91.64
IMNTPU-3 (Translation)	**92.16**	91.64

Performance- English

Run	Dev Set F1-Score(%)	Test Set F1-Score(%)
IMNTPU-1(Baseline)	87.13	88.39
IMNTPU-2(Double Redaction)	88.82	**89.86**

Conclusions

- IMNTPU Submitted Three runs for NTCIR-16 FinNUM3
  - IMNTPU1- XLM-RoBERTa Baseline Model
  - IMNTPU2- Double Redaction
  - IMNTPU3- Transaltion
- The performance with **data augmentation** method (Double Redaction) in **English** dataset is **superior** than without data augmentation.

Contribution

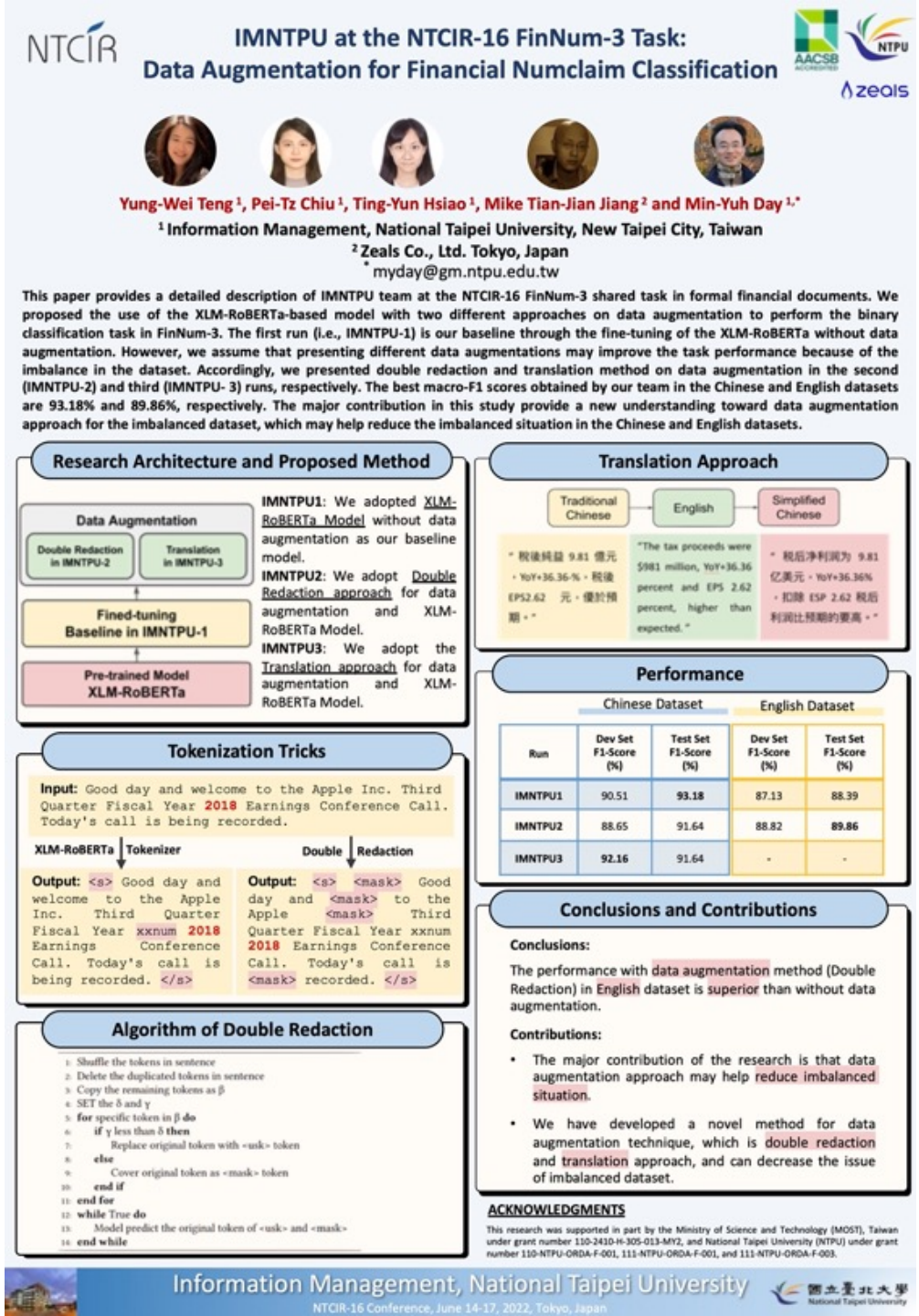
- The major contribution of the research is that data augmentation approach may help **reduce imbalanced situation**.
- We have developed a novel method for data augmentation technique, which is **double redaction** and **translation** approach, and can decrease the issue of imbalanced dataset.

NTCIR

2022

NTCIR-16

Best Poster Presentation Award



NTCIR

IMNTPU at the NTCIR-16 FinNum-3 Task: Data Augmentation for Financial Numclaim Classification

Yung-Wei Teng [1], Pei-Tz Chiu [1], Ting-Yun Hsiao [1], Mike Tian-Jian Jiang [2] and Min-Yuh Day [1,*]

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

[2] Zeals Co., Ltd. Tokyo, Japan

* myday@gm.ntpu.edu.tw

This paper provides a detailed description of IMNTPU team at the NTCIR-16 FinNum-3 shared task in formal financial documents. We proposed the use of the XLM-RoBERTa-based model with two different approaches on data augmentation to perform the binary classification task in FinNum-3. The first run (i.e., IMNTPU-1) is our baseline through the fine-tuning of the XLM-RoBERTa without data augmentation. However, we assume that presenting different data augmentations may improve the task performance because of the imbalance in the dataset. Accordingly, we presented double redaction and translation method on data augmentation in the second (IMNTPU-2) and third (IMNTPU- 3) runs, respectively. The best macro-F1 scores obtained by our team in the Chinese and English datasets are 93.18% and 89.86%, respectively. The major contribution in this study provide a new understanding toward data augmentation approach for the imbalanced dataset, which may help reduce the imbalanced situation in the Chinese and English datasets.

Research Architecture and Proposed Method

Data Augmentation
- Double Redaction in IMNTPU-2
- Translation in IMNTPU-3

Fined-tuning Baseline in IMNTPU-1

Pre-trained Model XLM-RoBERTa

IMNTPU1: We adopted XLM-RoBERTa Model without data augmentation as our baseline model.

IMNTPU2: We adopt Double Redaction approach for data augmentation and XLM-RoBERTa Model.

IMNTPU3: We adopt the Translation approach for data augmentation and XLM-RoBERTa Model.

Translation Approach

Traditional Chinese → English → Simplified Chinese

"税後純益 9.81 億元，YoY+36.36%，税後 EPS2.62 元，優於預期。"

"The tax proceeds were $981 million, YoY+36.36 percent and EPS 2.62 percent, higher than expected."

"税后净利润为 9.81 亿美元，YoY+36.36%，扣除 ESP 2.62 税后利润比预期的要高。"

Tokenization Tricks

Input: Good day and welcome to the Apple Inc. Third Quarter Fiscal Year 2018 Earnings Conference Call. Today's call is being recorded.

XLM-RoBERTa Tokenizer

Output: <s> Good day and welcome to the Apple Inc. Third Quarter Fiscal Year xxnum 2018 Earnings Conference Call. Today's call is being recorded. </s>

Double Redaction

Output: <s> <mask> Good day and <mask> to the Apple <mask> Third Quarter Fiscal Year xxnum 2018 Earnings Conference Call. Today's call is <mask> recorded. </s>

Performance

Run	Chinese Dataset Dev Set F1-Score (%)	Chinese Dataset Test Set F1-Score (%)	English Dataset Dev Set F1-Score (%)	English Dataset Test Set F1-Score (%)
IMNTPU1	90.51	**93.18**	87.13	88.39
IMNTPU2	88.65	91.64	88.82	**89.86**
IMNTPU3	**92.16**	91.64	-	-

Algorithm of Double Redaction

```
Shuffle the tokens in sentence
Delete the duplicated tokens in sentence
Copy the remaining tokens as β
SET the δ and γ
for specific token in β do
    if γ less than δ then
        Replace original token with «usk» token
    else
        Cover original token as «mask» token
    end if
end for
while True do
    Model predict the original token of «usk» and «mask»
end while
```

Conclusions and Contributions

Conclusions:

The performance with data augmentation method (Double Redaction) in English dataset is superior than without data augmentation.

Contributions:

- The major contribution of the research is that data augmentation approach may help reduce imbalanced situation.
- We have developed a novel method for data augmentation technique, which is double redaction and translation approach, and can decrease the issue of imbalanced dataset.

ACKNOWLEDGMENTS

This research was supported in part by the Ministry of Science and Technology (MOST), Taiwan under grant number 110-2410-H-305-013-MY2, and National Taipei University (NTPU) under grant number 110-NTPU-ORDA-F-001, 111-NTPU-ORDA-F-001, and 111-NTPU-ORDA-F-003.

Summary

- **NTCIR Evaluation of Information Access Technologies**
  - **NTCIR 16 (2021-2022)**
  - **NTCIR 17 (2022-2023)**
- **Dialogue System**
- **Financial Technology**

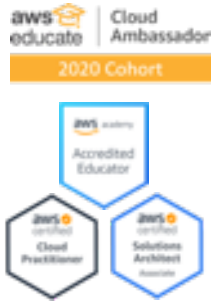
Acknowledgments: Research Projects

- 計畫主持人，應用 AI 技術建構加密貨幣反洗錢知識圖譜：少樣本學習模型 (Applying AI technology to construct knowledge graphs of cryptocurrency anti-money laundering: a few-shot learning model)，
  - 科技部 (人文司 - 商事財經法)，110-2410-H-305-013-MY2，2021/08/01~2023/07/31 [核定經費 (新台幣)：1,022,000]
- 子計畫共同主持人，深化企業永續-由人工智慧、財務與策略觀點打造企業永續績效 (Deepen Corporate Sustainability: Enhance the Performance of Corporate Sustainability from AI, Financial, and Strategic Perspectives)：子計畫二：人工智慧企業永續評鑑與跨語言永續績效報告書生成式模型 (AI for Corporate Sustainability Assessment and Cross Language Corporate Sustainability Reports Generative Model)
  - 國立臺北大學，111-NTPU_ORDA-F-001，2022/01/01~2022/12/31 [經費總額 (新台幣)： 3,228,500]
- 子計畫主持人，人工智慧方法分析企業科技創新導入- 專利文字分析與影像分析應用 (Artificial intelligence methods applied for analyzing the introduction of technological innovation: Patent text analysis and image analysis)：子計畫三：應用人工智慧於專利文本分析金融科技知識圖譜 (Artificial Intelligence for FinTech Knowledge Graph from Patent Textual Analytics)
  - 國立臺北大學，111-NTPU_ORDA-F-003，2022/01/01~2022/12/31 [經費總額 (新台幣)： 1,291,950]
- 子計畫共同主持人，企業永續動機、價值攸關性與人工智慧於企業永續績效評比之應用 (Corporate Sustainability: Motivations, Value Relevance, and the Application of AI in the Assessment)：子計畫二：人工智慧 AI 於企業永續評比之應用 (An application of artificial intelligence (AI) in the corporate sustainability assessment)
  - 國立臺北大學，110-NTPU_ORDA-F-001，2021/01/01~2021/12/31 [經費總額 (新台幣)： 3,240,000]

Q & A

NTCIR Evaluation of Dialogue System and Financial Technology

Host: 吳怡瑾 教授 (Prof. Nancy I-Chin Wu)
Graduate Institute of Library & Information Studies, National Taiwan Normal University (NTNU)
Time: 13:30-15:30, September 16, 2022 (Friday)
Place: Graduate Institute of Library & Information Studies, NTNU



戴敏育 副教授

Min-Yuh Day, Ph.D, Associate Professor

國立臺北大學 資訊管理研究所

Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

2022-09-16

References

- Takehiro Yamamoto and Zhicheng Dou (2022), "Overview of NTCIR-16", in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16), Tokyo Japan, June 14-17, 2022.
- Sijie Tao and Tetsuya Sakai (2022), "Overview of the NTCIR-16 Dialogue Evaluation (DialEval-2) Task", in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16), Tokyo Japan, June 14-17, 2022.
- Chung-Chi Chen, Hen-Hsen Huang, Yu-Lieh Huang, Hiroya Takamura and Hsin-Hsi Chen (2022), "Overview of the NTCIR-16 FinNum-3 Task: Investor's and Manager's Fine-grained Claim Detection", in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16), Tokyo Japan, June 14-17, 2022.
- Ting-Yun Hsiao, Yung-Wei Teng, Pei-Tz Chiu, Mike Tian-Jian Jiang, and Min-Yuh Day (2022), "IMNTPU Dialogue System Evaluation at the NTCIR-16 DialEval-2 Dialogue Quality and Nugget Detection", in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16), Tokyo Japan, June 14-17, 2022.
- Yung-Wei Teng, Pei-Tz Chiu, Ting-Yun Hsiao, Mike Tian-Jian Jiang, and Min-Yuh Day (2022), "IMNTPU at the NTCIR-16 FinNum-3 Task: Data augmentation for financial Numclaim classification", in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16), Tokyo Japan, June 14-17, 2022.