

# CCL23-Eval 任务6总结报告: 电信网络诈骗案件分类

孙承杰<sup>1\*</sup>, 纪杰<sup>1</sup>, 尚伯乐<sup>2</sup>, 刘秉权<sup>1</sup>

<sup>1</sup>哈尔滨工业大学, 哈尔滨, 150001

<sup>2</sup>哈尔滨市公安局香坊分局, 哈尔滨, 150001

sunchengjie@hit.edu.cn, jijie@insun.hit.edu.cn

shangboyue@163.com, liubq@hit.edu.cn

## 摘要

近年来, 电信网络诈骗形势较为严峻, 自动化案件分类有助于打击犯罪。本文介绍了任务相关的分类体系, 其次从数据集、任务介绍、比赛结果等方面介绍并展示了本次评测任务的相关信息。本次任务共有60支参赛队伍报名, 最终有34支队伍提交结果, 其中有15支队伍得分超过 baseline, 最高得分为0.8660, 高于 baseline 1.6%。根据结果分析, 大部分队伍均采用了 BERT 类模型。

**关键词:** 电信网络诈骗; 文本分类; BERT

## Overview of CCL23-Eval Task 6: Telecom Network Fraud Case Classification

Chengjie Sun<sup>1\*</sup>, Jie Ji<sup>1</sup>, Boyue Shang<sup>2</sup>, Bingquan Liu<sup>1</sup>

<sup>1</sup>Harbin Institute of Technology, Harbin, 150001

<sup>2</sup>Harbin Public Security Bureau Xiangfang Branch, Harbin, 150001

sunchengjie@hit.edu.cn, jijie@insun.hit.edu.cn

shangboyue@163.com, liubq@hit.edu.cn

## Abstract

In recent years, the situation of telecom network fraud has been severe, and automated case classification can help fight crime. This article introduces the task-related classification system, and then introduces and displays the relevant information of this evaluation task from the aspects of data sets, task introduction, and competition results. A total of 60 participating teams signed up for this task, and finally 34 teams submitted results, of which 15 teams scored more than baseline, the highest score was 0.8660, which was 1.6% higher than baseline. According to the analysis of the results, most of the teams have adopted the BERT-like model.

**Keywords:** Telecom Network Fraud, Text Classification, BERT

## 1 引言

随着互联网技术的发展, 一些新型的互联网犯罪行为不断涌现, 其中电信网络诈骗犯罪尤为突出。近年来, 电信网络诈骗犯罪形势严峻, 已经成为当前发案最高、损失最大、群众反映最强烈的突出犯罪 (刘玲玲 et al., 2023)。电信网络诈骗犯罪行为, 给国家和人民都带来了不小

的损失，不仅有物质上的损失，更有精神上的挫折，对于政府公信力、社会信任等都造成了严重的威胁，打击电信网络诈骗犯罪刻不容缓。2022年12月1日起，《中华人民共和国反电信网络诈骗法》正式施行(张维炜, 2022)，这表明了国家欲治理当前电信网络诈骗乱象的决心。

针对电信网络诈骗案件进行分类，是有效帮助公安部门针对电信网络诈骗犯罪行为进行治理、防范、宣传等措施的关键(王洁, 2019)。目前在电信网络诈骗案件的处理过程中，需要将案件上传至公安部门的反诈大数据平台并进行分类，而当前的分类工作主要依赖于人工进行标注，因此利用深度学习以及自然语言处理技术实现案件的自动分类，可以有效地降低人工成本，提高分类工作效率。

文本分类是自然语言处理领域的一个基础任务。在自然语言处理技术发展早期，基于统计模型的文本分类的方法占据主流，例如朴素贝叶斯、K近邻、SVM等机器学习模型。随着GPU计算能力的提高，基于深度学习的文本分类方法逐渐受到关注，例如TextCNN(Kim, 2014)、FastText(Joulin et al., 2017)、BERT(Devlin et al., 2019)等深度学习模型。案件分类主要是依据案件文本进行分类，其中案件文本为案件过程的描述性文本，具体来说，即对案件发生的时间、地点、经过进行了概括性描述的文本，因而案件分类可以看作是一个文本分类任务。在电信网络诈骗案件文本分类任务上，相关研究较少，目前还存在空白。因此，本任务的提出也是为了推进这一领域的研究进展。

分类任务需要按照一定的分类的体系或标准进行，目前电信网络诈骗的分类体系较多。宋兵(2019)按照法益以及受害者受骗领域的不同为标准，将电信网络诈骗分为金融理财型、消费购物型、虚假冒充型等3个类别。孙高峰(2020)按照受害人的心理，将案件分为由贪财、恐惧、好奇等3个心理类别。葛俊峰(2019)同样从受害人心理角度出发，将案件分为利诱、色诱、情诱、信诱、威逼等5个类别。2019年2月19日，公安部官方微博“中国警方在线”发布了最全60种典型电信网络诈骗手段，分为仿冒身份诈骗、购物类诈骗等8个类别。2022年8月1日，中国司法大数据研究院正式对外发布《涉信息网络犯罪特点和趋势(2017.1-2021.12)司法大数据专题报告》，报告中将电信网络诈骗分为贷款、冒充、招聘等7个类别。

本次评测任务采取我国公安部门的反诈平台现行的分类体系，具体来说有冒充电商物流客服类、虚假征信类等13个类别，由于“其他类型诈骗”数据较少且特征复杂，本次任务采用了除该类别以外的12个类别，具体说明可参考第2节。

本文组织结构如下：第一部分分析了电信网络诈骗的形势及案件分类的意义，同时介绍了现有的分类体系。第二部分从数据处理、数据样例、数据分布等三个方面，对数据集进行了介绍。第三部分介绍了任务的流程、评价标准、baseline等。第四部分主要展示了本次任务的比赛结果。第五部分对比赛结果及任务中模型采用情况进行了分析。第六部分对本次任务进行总结及未来展望。

## 2 数据集

### 2.1 数据处理

**数据采集** 数据由公安部门反诈大数据平台导出，每一条数据包含案件文本和类别标注，其中案件文本内容为案情简述，即关于案件经过的描述性文本，具体示例可参考下面的2.2。

**数据清洗** 从反诈大数据平台共计导出13个类别的数据，去除了“其他类型诈骗”类别，因此本次任务数据集共有12个类别。

**脱敏处理** 为防止对受害者造成二次伤害以及防止诈骗信息产生二次传播，去除了数据中受害者的隐私信息及诈骗分子的不良信息，具体来说，包括案件文本中的姓名、出生日期、地址、涉案网址、各类社交账号以及银行卡号码等信息。

**分类依据** 类别体系来源于反诈大数据平台的分类标准，主要依据受害人的法益及犯罪分子的手法进行分类，例如冒充淘宝客服谎称快递丢失的，分为冒充电商物流客服类；冒充公安、检察院、法院人员行骗的，分为冒充公检法及政府机关类；谎称可以帮助消除不良贷款记录的，分为虚假征信类等等，具体类别划分可参考下面的2.3。

### 2.2 数据样例

数据以json格式存储，每一条数据具有三个属性，分别为案件编号、案情描述、案件类别。样例如下：

```
{
```

```

    "案件编号": 28043,
    "案情描述": "事主（女，20岁，汉族，大专文化程度，未婚，现住址：）报称2022年8月27日13时43分许在口被嫌疑人冒充快递客服以申请理赔为由诈骗3634元人民币。对方通过电话（）与事主联系，对方自称是中通快递客服称事主的快递物件丢失现需要进行理赔，事主同意后对方方便让事主将资金转入对方所谓的“安全账号”内实施诈骗，事主通过网银的方式转账。事主使用的中国农业银行账号，嫌疑人信息：1、成都农村商业银行账号，收款人：；2、中国建设银行账号，收款人：。事主快递信息：中通快递，.现场勘查号：。",
    "案件类别": "冒充电商物流客服类"
}

```

### 2.3 数据分布

数据共有12个类别，各类别对应的样本数量如表1所示，相应的分布柱状图如图1所示。

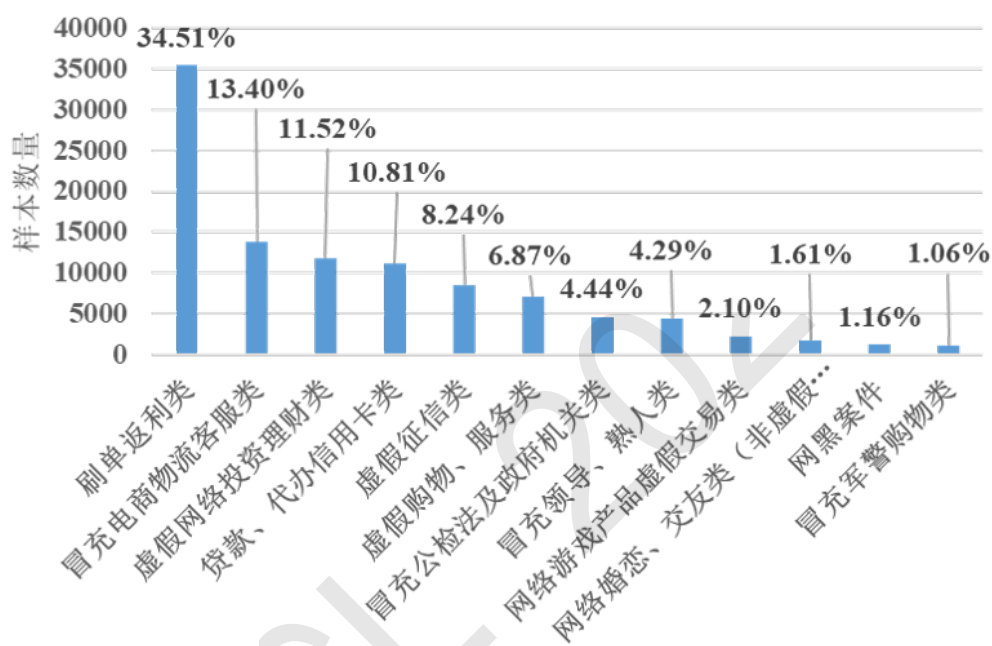


Figure 1: 数据分布柱形图

由图1可见，本次任务的数据集存在有数据不平衡情形，数据之间的分布差异较大，其中“刷单返利类”数量占据最多，接近总样本数量的三分之一。

数据集按8:1:1的比例，划分为训练集、测试集A、测试集B。具体数量如表2所示。本次任务仅采用了训练集及测试集A以作评测。

## 3 任务介绍

### 3.1 任务流程

本次任务目的是对给定案件描述文本进行分类。案件描述文本为对案件发生过程的描述性文本（经过脱敏处理，或称数据匿名化处理）。具体分类流程，如图2所示。

### 3.2 评价标准

评测性能时，本任务主要采用宏平均 F1 值作为评价标准，即对每一类计算 F1 值，最后取算术平均值，其计算方式如下：

$$Macro_{F1} = \frac{1}{n} \sum_{i=1}^n F1_i$$

类别名称	样本数量
刷单返利类	35459
冒充电商物流客服类	13772
虚假网络投资理财类	11836
贷款、代办信用卡类	11105
虚假征信类	8464
虚假购物、服务类	7058
冒充公检法及政府机关类	4563
冒充领导、熟人类	4407
网络游戏产品虚假交易类	2155
网络婚恋、交友类（非虚假网络投资理财类）	1654
网黑案件	1197
冒充军警购物类	1092
总计	102762

Table 1: 类别及对应样本数量

	样本数量
训练集	82210
测试集A	10276
测试集B	10276
总计	102762

Table 2: 数据划分情况

其中  $F1_i$  为第  $i$  类的 F1 值， $n$  为类别数，在本任务中  $n$  取12。

### 3.3 Baseline

本次任务采用经典的文本分类模型 TextCNN 和 BERT 作为 baseline。其中 TextCNN 为经典架构，主要由三层卷积层构成，卷积核的大小分别设置为2、3、4，池化层采用最大池化。采用的 BERT 模型为 base 版本，在 BERT 输出层后加入一层全连接层，结合微调方法训练。baseline 结果如表3所示。

Model	Macro Avg F1
TextCNN	0.8464
BERT	0.8503

Table 3: baseline 得分结果

## 4 比赛结果

本次评测共有60支参赛队伍报名。最终有34支参赛队伍提交结果。共有15支队伍得分超过 baseline。鉴于篇幅有限，仅展示超过 baseline 的队伍得分结果，如表4所示。

第一名使用了多个不同的预训练模型进行评测，包括 Chinese-RoBERTa-wwm-ext-large (Cui et al., 2021)、NEZHA-base-wwm (Wei et al., 2019) 和 Erlangshen-DeBERTa-v2 (Wang et al., 2022) 等系列模型，在这些模型上进行领域预训练 (Sun et al., 2019; Gururangan et al., 2020) 任务，同时结合了 FreeLB (Zhu et al., 2019) 的对抗训练方法，最终提交结果融合了 NEZHA 模型与参数量为320M、710M的 Erlangshen 模型。

第二名尝试了 BERT-base、BERT-wwm、BERT-wwm-ext、ERNIE (Sun et al., 2019) 和 Chinese-RoBERTa-wwm-ext 等多个预训练模型，根据实验结果，选择了 BERT-wwm 和 ERNIE 作为基底模型，针对数量较少的类别采取了伪标签 (Rizve et al., 2021) 的方法增加了训

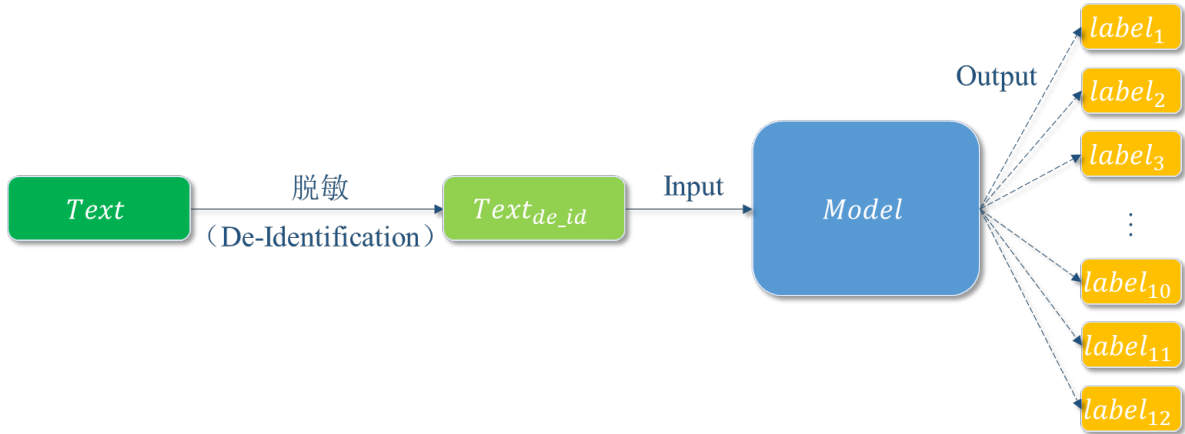


Figure 2: 任务流程示意图

Team Name	Macro Avg F1
CVTEDMer	0.866068
ZUT	0.862891
GDUT-Hyq	0.862466
DECEM	0.861416
NARI-DCloud	0.860952
KDSEC@IIE	0.860242
翼智团@TeleAI	0.860104
SCUT-SSE	0.859259
SUDA-GLnb	0.857642
KNODI-NLP	0.857474
BLCU-ST	0.855343
BIT	0.852768
HNU	0.852425
NENU-NLP	0.851454
CCNU	0.850499

Table 4: 部分队伍得分结果 (超过 baseline)

练集数量，最后提交结果融合了这两个基底模型。

第三名以 NEZHA 模型为基底模型，采用了 FGM (Miyato et al., 2016) 对抗训练方法增加扰动，并使用了指数移动平均策略更新网络权重。

第五名基于 NEZHA 模型，采用了层次分解方法延拓了位置编码向量，同时也采用了对抗训练策略增加模型的鲁棒性。

第六名采用 Chinese-BERT-wwm-ext 作为基底模型，通过对 CLS 位置进行动态加权平均增强向量的语义表征能力，同时采用 Multi-Sample Dropout 方法对输出进行多次 Dropout 以增强模型泛化能力，此外还通过回译 (Wei and Zou, 2019) 的方式，扩充了数量较少类别的样本规模，最后同样在训练过程中加入了 FGM 对抗训练方法。

第七名基于原型监督对比学习 (Wang et al., 2021) 思想，以 BERT-base、BERT-large 为基底模型，结合了领域预训练、FGM 对抗训练、R-Drop (Liang et al., 2021) 等方法，并针对混淆程度较高的类别使用后置分类模块进行二次补充判别。

第八名采用 Chinese-RoBERTa-wwm-ext、LERT (Cui et al., 2022)、MacBERT (Cui et al., 2020) 等作为基底模型，结合领域预训练方法，以增强模型语义理解能力，通过 FGM、AWP (Dong et al., 2020) 对抗训练和多次随机采样方法，以提高模型语义挖掘能力。此外，还将伪标签 (Lee, 2013) 样本添加到训练集中，进行了语义增强。最后结合了模型融合方法进行预测以获得结果。

本次任务还通过问卷统计了各队伍所采用的模型情况，结合已收集的技术报告，鉴于问卷采取自愿填写方式且存在无效回答，因此未能统计全部参赛队伍的情况，不完全统计结果如图3所示。

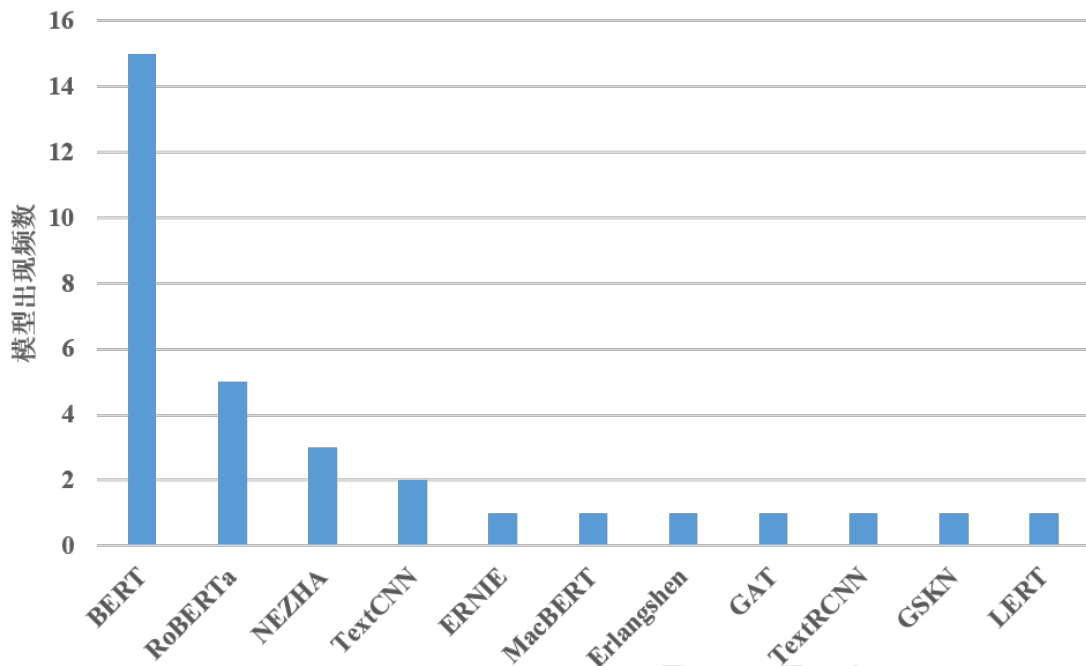


Figure 3: 模型使用情况（不完全）统计结果

## 5 结果分析

从本次任务采用模型来看，本次评测队伍基本采用的是 BERT 类的模型，典型的包括 BERT、RoBERTa 等，还有中文预训练模型 ERNIE、NEZHA 等。这表明在目前的文本分类任务上，主要还是以传统的预训练模型结合微调技术的方法为主，而当下较为火热的大模型，如 ChatGPT 等并未被采用进行本任务。当然，这可能跟本次任务未能提供相关支持有关。

从本次任务的结果来看，对抗训练、模型融合是使用最多且有效的方法。作为案件简述文本来说，其中噪声较多，对于分类并无帮助的文本内容较多，例如各类未涉及隐私的个人信息、时间、银行名称等，同时由于数据脱敏时采用的方式较为直接，文本中残留不少标点符号，因此对抗训练通过增加扰动，可能增强了模型对于这类信息的抗干扰性，从而提高了模型性能。此外，模型融合作为一般性策略，在各类评测任务中均有较好效果，本次任务结果也验证了这一点。

在数据不平衡方面，部分队伍通过伪标签或回译的方法取得了较好的结果。伪标签方法通过将无标签的测试集样本，经模型预测后标注伪标签，加入到训练集中，增加了训练集数量，从结果上来看，有效缓解了部分类别样本数量较少的问题。回译方法通过将数据样本翻译为另一语言甚至多个语言后，再翻译回原语言的方式，同样增加了不平衡类别的样本数量，从而提高了模型结果。

本次任务可看作是电信网络诈骗领域的文本分类任务，即特定领域文本的分类任务，因此部分队伍采用了领域预训练的方法，通过在本次任务的领域语料上进行继续预训练任务，使模型能够学习到领域中数据分布特征。本次任务结果也验证了该方法的有效性。

此外，少数队伍注意到了数据中存在易混淆的类别，并提出后置分类的方法，以解决该问题。该方法通过在模型预测所有类别后，针对易混淆类别进行二次补充判别，最终提高了易混淆类别识别效果以及模型整体的预测结果。易混淆类别的存在，根源于分类体系的定义不够明确清晰，类别之间划分边界较为模糊，从而导致类别混淆、类别重叠等问题。

## 6 结论及未来展望

本次任务中 BERT 类模型结合对抗训练方法取得较好结果，传统微调预训练方法仍占据主流，比赛结果中最高得分比 baseline 提升了 1.6%，达到 0.8660，表明该任务还存在较大的提升空间。未来将考虑进一步优化分类数据集，包括数据集中的噪声信息，以及一些错误分类的数据项，需要进行处理。此外当前分类体系下部分数据存在有分类重叠等问题，还需要进一步改进分类体系。

## 参考文献

- Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 427–431.
- Chen Zhu, Yu Cheng, Zhe Gan, Siqi Sun, Thomas Goldstein, and Jingjing Liu. 2019. Freelib: Enhanced adversarial training for language understanding. *arXiv preprint arXiv: 1909.11764*.
- Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune bert for text classification?. In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18*, 194–206.
- Dong-Hyun Lee. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, 896.
- Jason Wei, Kai Zou 2019. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 6382–6388.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.
- Junjie Wang, Yuxiang Zhang, Lin Zhang, Ping Yang, Xinyu Gao, Ziwei Wu, Xiaoqun Dong, Junqing He, Jianheng Zhuo, Qi Yang, Yongfeng Huang, Xiayu Li, Yanghan Wu, Junyu Lu, Xinyu Zhu, Weifeng Chen, Ting Han, Kunhao Pan, Rui Wang, Hao Wang, Xiaojun Wu, Zhongshen Zeng, Chongpei Chen, Ruyi Gan, and Jiaying Zhang 2022. Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence. *CoRR*, abs/2209.02970.
- Junqiu Wei, Xiaozhe Ren, Xiaoguang Li, Wenyong Huang, Yi Liao, Yasheng Wang, Jiashu Lin, Xin Jiang, Xiao Chen, and Qun Liu. 2019. Nezha: Neural contextualized representation for chinese language understanding. *arXiv preprint arXiv:1909.00204*.
- Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. 2021. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv preprint arXiv:2101.06329*.
- Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. 2021. Contrastive Learning based Hybrid Networks for Long-Tailed Image Classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 943–952.
- Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8342–8360.
- Takeru Miyato, Andrew M. Dai, and Ian Goodfellow. 2016. Adversarial training methods for semi-supervised text classification. *arXiv preprint arXiv:1605.07725*.
- Xiaobo Liang, Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, and Tie-Yan Liu. 2021. R-drop: Regularized dropout for neural networks. In *Advances in Neural Information Processing Systems*, 10890–10905.

- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 657–668.
- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. In *IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 29*, 3504–3514.
- Yiming Cui, Wanxiang Che, Shijin Wang, and Ting Liu. 2022. LERT: A Linguistically-motivated Pre-trained Language Model. *arXiv preprint arXiv:2211.05344*.
- Yinpeng Dong, Zhijie Deng, Tianyu Pang, Jun Zhu, and Hang Su. 2020. Adversarial distributional training for robust deep learning. In *Advances in Neural Information Processing Systems*, 8270–8283.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*.
- Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. ERNIE: Enhanced Representation through Knowledge Integration. *arXiv:1904.09223*.
- 葛俊峰. 2019. 深圳市电信网络诈骗特征与治理困境研究. 深圳大学.
- 刘玲玲, 毕梦瀛, 沈小晓. 2023. 多国出台措施打击电信网络诈骗. 人民日报, 2023-01-05(017).
- 宋兵. 2019. 电信网络诈骗犯罪的刑事立体防治. 青岛大学.
- 孙高峰. 2020. 电信网络诈骗犯罪现状与对策研究. 河北大学.
- 王洁. 2019. 电信网络诈骗犯罪的独特属性与治理路径. 中国人民公安大学学报(社会科学版), 35(04):1-10.
- 张维炜. 2022. 密织反诈“防护网” 压实“守门人”责任——反电信网络诈骗法正式实施. 中国人大, No.563(23): 33-34.