



计算机科学

COMPUTER SCIENCE

预训练语言模型的应用综述

孙凯丽, 罗旭东, 罗有容

引用本文

孙凯丽, 罗旭东, 罗有容. 预训练语言模型的应用综述[J]. 计算机科学, 2023, 50(1): 176-184.

SUN Kaili, LUO Xudong, Michael Y.LUO. Survey of Applications of Pretrained Language Models[J]. Computer Science, 2023, 50(1): 176-184.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

一种基于强化学习的口令猜解模型

Password Guessing Model Based on Reinforcement Learning

计算机科学, 2023, 50(1): 334-341. <https://doi.org/10.11896/jsjcx.211100001>

基于双向注意力机制和门控图卷积网络的文本分类方法

Text Classification Method Based on Bidirectional Attention and Gated Graph Convolutional Networks

计算机科学, 2023, 50(1): 221-228. <https://doi.org/10.11896/jsjcx.211100095>

利用异构图神经网络实现情绪-原因对的有效抽取

Utilizing Heterogeneous Graph Neural Network to Extract Emotion-Cause Pairs Effectively

计算机科学, 2023, 50(1): 205-212. <https://doi.org/10.11896/jsjcx.211100265>

融合注意力特征的无锚框视觉目标跟踪方法

AFTM:Anchor-free Object Tracking Method with Attention Features

计算机科学, 2023, 50(1): 138-146. <https://doi.org/10.11896/jsjcx.211000083>

亮度自调节的无监督图像去雾与低光图像增强算法研究

Study on Unsupervised Image Dehazing and Low-light Image Enhancement Algorithms Based on Luminance Adjustment

计算机科学, 2023, 50(1): 123-130. <https://doi.org/10.11896/jsjcx.211100058>

预训练语言模型的应用综述

孙凯丽^{1,2} 罗旭东^{1,2} 罗有容³

1 广西多源信息挖掘与安全重点实验室 广西 桂林 541004

2 广西师范大学计算机科学与工程学院/软件学院 广西 桂林 541004

3 剑桥大学伊曼纽尔学院 英格兰 剑桥 CB2 3AP

(Kaili.Sun@foxmail.com)

摘要 近年来,预训练语言模型发展迅速,将自然语言处理推到了一个全新的发展阶段。文中的综述旨在帮助研究人员了解强大的预训练语言模型在何处以及如何应用于自然语言处理。具体来讲,首先简要回顾了典型的预训练模型,包括单语言预训练模型、多语言预训练模型以及中文预训练模型;然后讨论了这些预训练模型对5个不同的自然语言处理任务的贡献,即信息提取、情感分析、问答系统、文本摘要和机器翻译;最后讨论了预训练模型的应用所面临的一些挑战。

关键词: 预训练语言模型;自然语言处理;深度学习;信息提取;情感分析;问答系统;文本摘要;机器翻译

中图法分类号 TP391

Survey of Applications of Pretrained Language Models

SUN Kaili^{1,2}, LUO Xudong^{1,2} and Michael Y. LUO³

1 Guangxi Key Lab of Multi-source Information Mining & Security, Guilin, Guangxi 541004, China

2 School of Computer Science and Engineering & School of Software Guangxi Normal University, Guilin, Guangxi 541004, China

3 Emmanuel College, Cambridge University, Cambridge, CB2 3AP, UK

Abstract In recent years, pretrained language models have developed rapidly, pushing natural language processing into a whole new stage of development. To help researchers understand where and how the powerful pretrained language models can be applied in natural language processing, this paper surveys the state-of-the-art of its application. Specifically, we first briefly review typical pretrained language models, including monolingual, multilingual and Chinese pretrained models. Then, we discuss these pretrained language models' contributions to five different natural language processing tasks: information extraction, sentiment analysis, question answering, text summarization, and machine translation. Finally, we discuss some challenges faced by the applications of pretrained language models.

Keywords Pretrained language model, Natural language process, Deep learning, Information extraction, Sentiment analysis, Question answering system, Text summarization, Machine translation

1 引言

在早期的自然语言处理(Natural Language Processing, NLP)任务中,人们通常利用各种深度神经网络^[1],使用低维密集向量来表示语言的句法或语义特征。然而,由于深度神经网络参数量大,并且在有限的数据集上进行训练的,往往会导致过度拟合。因此,Vaswanid等^[2]提出了深度学习模型Transformer,用于解决循环神经网络(Recurrent Neural Network, RNN)训练缓慢、效率低下和过度拟合的问题^[3]。同时,将Transformer与自我注意力机制相结合,实现了快速并行的效果。从此,预训练语言模型(Pretrained Language Model, PLM)进入了蓬勃发展的阶段。各种大规模的单语言

PLM(如BERT^[4]和GPT^[5])在各种NLP任务中获得了巨大成功,并催生了许多不同的改进模型和多语言PLM。

近两年来,PLM在NLP领域应用得非常广泛,原因是PLM可以将各种困难复杂的NLP任务简化为简单的微调问题。PLM通过自我监督学习在大规模文本语料库上进行预训练,预训练过程包括从大型语料库中学习常识(与特定的下游NLP任务无关),以防止模型在小数据上过度拟合^[6]。在预训练之后,PLM需要针对具体的下游任务进行微调。

由于PLM对NLP至关重要,为了帮助研究人员快速掌握PLM的应用现状,本文综述了PLM在NLP领域5个最热门的任务中的应用:信息提取(Information Extraction, IE)、情感分析(Sentiment Analysis, SA)、文本摘要(Text Summa-

到稿日期:2022-08-23 返修日期:2022-10-10

项目基金:广西多源信息挖掘与安全重点实验室系统性研究课题(22-A-01-02)

This work was supported by the Guangxi Key Lab of Multi-source Information Mining & Security(22-A-01-02).

通信作者:罗旭东(luoxd@mailbox.gxnu.edu.cn)

rization, TS)、问题回答(Question Answering, QA)和机器翻译(Machine Translation, MT)。此外,我们还讨论了当前PLM应用中的重要挑战。

其他研究人员也综述了PLM,但本文的综述与他们的不同。2020年,Qiu等^[7]对PLM在NLP领域的应用进行了调查,但只进行了简单的概述。2020年,Zaib等^[8]对PLM的对话式人工智能做了一个简短的综述,但未涉及PLM在其他领域的应用。2021年,Luo等^[9]调查了PLM在跨语言SA中的应用。2021年,Kalyan等^[10]回顾了基于转换的生物医学PLM。然而,他们的PLM只涉及生物医学领域,这些综述都没有包含最新的进展,且所做的综述单一或者不够详尽。本文不仅讨论了PLM在NLP领域的最新应用,而且详细地介

绍了这个主题,涵盖了信息提取、情感分析、自动问答、文本摘要和机器翻译5个NLP热门领域。

本文第2节简要介绍了一些典型的PLM,包括国内外的单语和多语PLM;第3节总结了一些基于PLM的信息提取方法;第4节回顾了一些基于PLM的情感分析技术;第5节考察了用于问答系统的PLM模型;第6节讨论了基于PLM的文本摘要方法;第7节评论了一些关于PLM的机器翻译的研究;第8节讨论了未来工作的挑战;最后总结全文。

2 预训练模型

本节将简要介绍一些典型的国内外的单语和多语PLM,如表1所列。

表1 一些典型的PLM的简要总结
Table1 Brief summary of some typical PLM

PLM	来源	特点	优点	缺点	
BERT家族	BERT	文献[4]	使用MLM和NSP进行双向联合训练	在NLP的11项基本任务上取得了重大的进展	预训练阶段mask掉的单词之间有的具有相关性,但BERT未考虑到这一点
	AIBERT	文献[11]	可以跨层共享参数	参数量大大减少	计算成本高
	ELECTRA	文献[12]	将BERT中的MLM替换为替换令牌检测	解决了BERT在MASK的预训练阶段和微调阶段的不一致问题	未同时充分考虑模型的计算效率和绝对性能
	XLNet	文献[13]	使用ARLM并吸收了Transformer-XL的思想	在20个NLP任务上超越了BERT,并在18个NLP任务上取得了最先进的成果	不能很好地应用于视觉和强化学习领域
单语言预训练模型	GPT	文献[14]	无监督的预训练和有监督的微调	在自然语言推理、分类和问题回答方面优于以前的模型	GPT是单向模型,无法很好地利用上下文信息
	GPT-2	文献[15]	训练数据在数量、质量和广度上都有了很大的提升,最大规模的GPT-2的网络参数达到了48层	无需特定领域的培训即可回答问题、总结文本和翻译文本,对长文本(长距离依赖)表现良好	在许多实际任务中,GPT-2的性能要差于随机性
	GPT-3	文献[16]	有1750亿个参数	在文本生成方面,与人工生成的文本几乎没有区别	没有很好地遵循用户的真实意图,并且经常输出不真实、有害或反应迟钝的情感文本
中文PLM	ERNIE	文献[17]	对训练数据中的语法结构、词汇结构和语义信息进行统一建模	在对远程监督数据进行去噪和对有限数据进行微调方面的性能优于BERT	与BERT相比,ERNIE在小数据集上变得更加不稳定
	TinyBERT	文献[18]	使用一种新的知识蒸馏方法在预训练和特定任务学习阶段进行变压器蒸馏	大小仅为BERT的13.3%,但其计算速度是BERT的9.4倍	未能实现高效地将知识从更广泛和更深入的教师模型(例如BERTLARGE)转移到学生模型(TinyBERT)
	BERT-wwm	文献[19]	在预训练阶段采用全词覆盖方法	在正式文本和长文本建模任务上表现优异	在非正式文本建模任务上表现较差
多语言预训练模型	mBERT	文献[4]	在104种语言的单语语料库上进行预训练	跨语言转移的能力优于当前其他模型	面对不同语言时,存在信息不兼容问题
	XLM	文献[20]	单语数据集上的无监督学习和平行语料库上的监督学习	在XNLI数据集上,XLM以4.9%的绝对准确率领先于其他最先进的模型	严重依赖于跨语言迁移能力,不能对低资源语言进行建模
	XLM-R	文献[21]	使用超过2TB的Common-Crawl过滤数据对100种语言进行了训练	首次展示了在不牺牲全语言性能的情况下进行多语言建模的可能性	全世界存在7100多种语言,但只在100种语言上进行了训练

2.1 单语言预训练模型

2018年底,Devlin推出了单语言预训练模型BERT,它使用掩码语言模型(Mask Language Model, MLM)和下句预测(Next Sentence Prediction, NSP)进行深度双向联合训练。BERT的预训练和下游的特定任务训练之间的区别只是顶部输出层BERT在NLP的11个基本任务上取得了卓越的性能。此后,研究人员提出了许多BERT的改进版本。例如, RoBERTa^[22]使用更广泛的数据集,将静态掩码改为动态掩码,并取消了NSP任务。AIBERT^[11]可以跨层共享参数,大

大减少了参数。ELECTRA^[12]用替换令牌检测(Replaced Token Detection, RTD)取代了BERT中的MLM,解决了隐藏(MASK)的预训练和微调阶段的不一致问题。

XLNet^[13]与BERT相比在以下两个方面进行了改进: 1)BERT在MLM阶段所屏蔽的一些词是独立的,然而有些词之间是相互关联的,BERT却忽略了词之间的关联性。XLNet使用自动回归语言模型代替BERT中的自动编码器语言模型,消除了屏蔽标记的负面影响。在预测单词时,XLNet采用双流自我关注机制,同时考虑了一个单词的位置和

上下文信息;2)XLNet 采用了 Transformer-XL 架构。BERT 在处理超长序列时容易出现信息损失,但 Transformer-XL 采用了分段循环机制,可以有效处理超长序列的问题。XLNet 在 20 个 NLP 任务上的表现优于 BERT,并在其中 18 个任务上取得了最先进的效果。

2018 年,Radford 等^[14]研发了 GPT 模型,其预训练由两个阶段组成。1)无监督的预训练:在大型文本语料库上学习一个高容量的语言模型。2)有监督的微调:使模型适应有标签数据的判别性任务。2019 年,Radford 等^[15]研发了 GPT-2,它仍然使用 GPT 的单向转化器模式,但训练数据在数量、质量和广度上都有极大的提高,网络参数也有所增加,无需特定领域的训练就可以进行特定领域的问题回答、文本摘要和文本翻译。2020 年,Brown 等^[16]进一步研发出了 GPT-3,其拥有 1 750 亿个参数。GPT-3 在许多 NLP 任务上表现出色,包括机器翻译、问题回答和文本摘要。

在英语预训练模型繁荣之后,中文预训练模型也得到了巨大和快速的发展。2019 年,百度发布了中文预训练模型 ERNIE^[17]。它通过统一建模语法结构、词汇结构和语义信息,大大提高了通用语义表示能力。同年,华为提出了 TinyBERT^[18],它的规模只有 BERT 的 13.3%,但其计算速度是 BERT 的 9.4 倍。哈尔滨工业大学和科大讯飞联合实验室在 BERT 的基础上,研发了中文预训练模型 BERT-wwm^[19]。BERT-wwm 是在谷歌发布的 BERT-base 中文版本的基础上研发的,在预训练阶段采用了全词覆盖方法。与 BERT 和 ERNIE 以及其他扩展模型的对比实验,证实了应用了全词覆盖技术的 BERT-wwm 更有效^[23]。

2.2 多语言预训练模型

除了这些大型的单语言 PLM,研究人员还开发了多语言 PLM。2019 年,Devlin 等研发了多语言的 BERT (multilingual BERT, mBERT)^[4],它可以为 104 种语言的各种 NLP 任务提供一个上下文表示。mBERT 采用 MLM 方法,在 104 种语言的单语语料库上进行了预训练,并采用正则化策略进行了改进,在跨语言的 NLP 任务中表现良好。

2019 年,Facebook 人工智能团队发布了跨语言预训练模型 XLM^[20]。他们使用两种方法进行跨语言建模:1)依靠单语数据集进行无监督学习;2)使用平行语料进行监督学习。在 XNLI 数据集上,他们的实验表明 XLM 以 4.9% 的绝对准确率超越了最先进的基线模型。

XLM 的改进模型 XLM-R^[21]是基于 XLM 的想法,并使用超过 2TB 的 CommonCrawl 过滤数据对 100 种语言进行训练。Facebook 人工智能团队通过调整模型的参数来解决多语言诅咒的问题,首次展示了在不牺牲全语言性能的情况下进行多语言建模的可能性。在 XNLI 数据集上,XLM-R 以 82% 的平均准确率达到目前的技术水平。更值得注意的是,XLM-R 在低资源语言上的表现特别好,准确率达到 23%,超过了最先进的基线模型。

3 信息抽取

本节将总结一些基于 PLM 的信息提取 (Information Extraction, IE) 方法。IE 的任务是从这些非结构化文本 (现实

生活中的大多数文档都是非结构化的)中识别和提取各种实体以及这些实体之间的关系,并将它们汇集成统一的形式^[24]。IE 可以帮助人们从许多文档中快速提取特定需求的事实信息。命名实体识别 (Named Entity Recognition, NER) 和关系提取 (Relation Extraction, RE) 是信息抽取的两个基本子任务。NER 旨在识别文本中具有特定含义的实体 (如手机和电脑),并将它们归入相应的实体类别。RE 是 NER 之后的步骤,其主要目的是根据上下文确定一对特定实体之间的关系。通常情况下,一段文本甚至一个句子中存在不止一对实体,这意味着不止一种关系,这就要求尽可能多地、正确地提取文本中的关系实体对。

3.1 命名体识别

由于许多词具有多义性,中文词的表示对于 NER 方法的准确性至关重要。为此,Cai^[25]提出了一个基于 BERT 的中文 NER 模型,即 BERT-BiGRU-CRF。该模型首先利用 BERT 根据词的上下文生成一个更准确的表示,然后将其输入 BiGRU-CRF 进行训练。该模型在微软命名体识别语料库中的表现优于最先进的 Lattice LSTM 模型。Chen 等^[26]不仅在 BiGRU-CRF 模型中整合了 BERT,还引入了一个记忆网络模块。BERT 可以更好地获取语义表示,记忆网络模块可以对更大范围的语义特征进行编码。他们在 CLUENER 2020 数据集上进行的实验表明,融入了 BERT 和记忆网络后的模型命名体的识别效果有所提升。Zhang 等^[27]将 NER 视为一个包含查询、语境和答案的三要素,为每个实体构建一个查询语句,并用 BERT 对查询和语境进行编码,在特征映射后与旋转位置嵌入相融合。他们用双峰注意力对文本的每个子序列进行评分后,再用平衡的 softmax 判断子序列是否为实体,通过多次实验证实了他们的方法在识别法律文件中的嵌套和非嵌套实体任务上效果良好。

为了提高中国电力设备故障知识图谱中 NER 的准确度,Meng 等^[28]提出了一个 NER 模型,称为 BERT-BiLSTM-CRF。首先,它使用 BERT 来提取句子层面的隐含特征,对中国电力设备故障文档进行初步 NER。然后,它使用 BiLSTM 来提取句子的局部特征。最后,他们通过条件随机场 (Conditional Random Field, CRF) 模型得到了最优的全局解决方案。与 LSTM, BiLSTM 和 BiLSTM-CRF 相比,他们的实验表明,加入 BERT 可以显著提高命名体识别的效率和准确性。

Moon 等^[29]提出了一个基于 mBERT 的常见语言 NER 模型。该模型无须训练即可在新语言上运行,并在多语言和单语言数据集上均取得了最先进的结果。Chen 等^[30]提出了一个基于 XLM-R 的低资源 NER 模型,称为 LRLFiT,该模型解决了在低资源语言中由于缺乏注释数据而造成的较难识别命名体的问题。在对低资源数据集进行数据扩充后,他们基于自我注意机制,对 XLM-R 进行了领域微调和特定任务微调。经过微调,XLM-R 可以输出更多相关的语义和句法信息。在维吾尔语和匈牙利语数据集上,该模型的 F1 值分别为 75.35% 和 92.45%,优于其他基线模型。

为了提高低资源语言中生物医学 NER 的准确性,Boudjellal 等^[31]提出了一个基于 BERT 的 NER 模型,称为 ABioNER。他们首先在一个普通领域的阿拉伯语语料库和

一个生物医学领域的阿拉伯语文献语料库上对阿拉伯语 BERT 进行了预训练,然后使用 NVIDIA GPU 对阿拉伯语 BERT 进行了微调。他们的实验表明,在小规模的生物医学数据上对单语 BERT 模型进行微调,对生物医学领域的 NER 是有效的。

在对大规模生物医学文本库进行 BERT 预训练后,Agrawal 等^[32]使用迁移学习对 BERT 进行了微调,以解决生物医学领域的嵌套实体识别问题。他们的实验表明,在 GENIA 数据集上,BERT 的表现优于其他基线模型。

3.2 关系抽取

许多研究人员将 NER 和 RE 视为独立的任务,但它们之间存在很强的联系,因为 NER 的错误可能会传播到 RE。为了解决这个问题,Han 等^[33]提出了一个基于 BERT 的 RE 模型。该模型通过实体屏蔽方法处理文档中的实体信息,然后使用 BERT 处理整个文档,以识别文档中每个实体之间的信息。他们的实验表明,在 DocRED 手工注释数据集上,与未使用任何 PLM 的最先进模型相比,他们的模型在 F1 值上取得了 6% 的提高。

Qiao 等^[34]构建了一个名为 BERT-BILSTM-LSTM 的农业实体-关系联合提取模型。他们在联合学习模型 LSTM-LSTM-Bias 中加入了微调的 BERT,从而弥补了联合学习模型不能处理多义词的不足。他们的实验表明,与单独的联合学习模型相比,加入 BERT 后的模型的 F1 值达到了显著提高。

为了提高 RE 对句子语义信息的捕获能力,Xu 等^[35]构建了一个基于 BERT 的关系抽取模型,称为 BERT-GMAN。他们利用 BERT 对句子和约束信息进行编码,形成动态词向量,增强了实体的上下文语义表示,减少了噪音干扰。在 SemEval-2010 任务 8 数据集上,BERT-GMAN 的 F1 值达到了 90.25%。

4 情感分析

本节将回顾一些基于 PLM 的情感分析(Sentiment Analysis,SA)方法(NLP 中最活跃的话题之一)。情感分析的主要任务是分析和处理带有鲜明态度的主观文本,以判断文本所表达的态度极性^[36]。

4.1 不同水平的情感分析

文档级情感分析将文档作为整体对象,并识别整个文档表达的是积极的还是消极的情绪^[37]。当使用 PLM 进行文档级情感分析时,文档的长度通常超过 PLM 的最大可接受长度。因此,PLM 将不得不截断该文件。然而,截断的过程可能会降低情感识别的准确性。

为了解决这个问题,Kong 等^[38]将文档的一部分作为一个区域,用 BERT 来学习输入文档中的句子级特征。同时,他们采取了自适应微调策略,对 BERT 的各层进行了动态微调,不需要对输入文本的所有层进行微调,从而减少了计算资源的消耗。

方面级情感分析是情感分析中的一个细粒度任务,与文档级情感分析相比,方面级情感分析可以更准确地分析和判断文本中的情感。对于方面级情感分析,Liao^[39]使用预训练

模型 RoBERTa 进行文本和方面标记的特征提取,并使用文档关注来加强对全文的理解。此外,他们使用交叉关注机制,对文本中与特定方面类别最相关的部分给予更多关注,使 RoBERTa 能够更准确地提取文本所表达的情感。

4.2 多语言的情感分析

Chen 等^[40]使用 mBERT 获得文本的语义特征,在检测德拉维德语系的文本中的攻击性语言时,用获取的语义特征来构建文本映射矩阵。他们通过 TextCNN 进行卷积运算后得到文本语义特征矩阵,并进一步利用池化算法降低维度得到文本语义特征向量。

mBERT 存在着不同语言信息不兼容的问题,但 XLM-R 可以在不同语言中一起进行预训练,具有良好的跨语言性能。因此,Jayanthi 等^[41]将 mBERT 与 XLM-R 整合在一起,对由马拉雅拉姆语、泰米尔语和卡纳达语 3 种语言组成的文本进行情感分析,取得了良好的效果。

虽然 mBERT 和 XLM-R 学会了许多语言,但它们还没有在许多资源匮乏的语言中接受过训练。因此,Xia 等^[42]以 mBERT 和 XLM-R 为基本模型,结合元学习和迁移学习方法,将辅助语言转换为目标语言,并使其表征更加接近,以促进迁移效果。

4.3 特定领域的情感分析

PLM 通常是在通用的语料库上训练的,因此将 PLM 应用于特定领域的情感分析非常具有挑战性。金融、生物医学和其他领域经常用到各种专业术语,但一般的语料库很少涵盖各种专业术语,而且对大量专业术语进行标记的成本特别高,因此一般的语言模型在某些特定领域是无效的。为了更好地将 PLM 应用于金融领域,Araci^[43]在一个大规模的金融语料库上训练了 BERT,并在 [CLS] 标记的最后一个隐藏状态之后添加了一个密集层,使得 BERT 能够在金融领域更好地进行情感分析。类似地,Cao 等^[44]使用 BERT 提取源域和目标域中的文本特征,并将其映射到共享的特征空间。这种方法可以解决下游任务,以更好地实现跨领域的情感分析。

5 问答系统

本节将考察一些基于 PLM 的问答系统(Question Answering System,QAS)的方法。问答系统的目的是理解人类用户的问题或查询,然后用自然语言^[45]回答他们,主要有两种类型。1)限定域的问答系统:在这种问答系统中,问答环节只针对一个特定的领域或有限的内容范围。比如,只回答医学领域的查询,或只回答关于 COVID-19 响应措施的查询。2)开放域的问答系统:开放域问答需要从不同领域的众多文档中寻找答案,需要大量的常识性知识和语义字典来正确处理文档。

5.1 限定域问答

Oniani 等^[46]用 COVID-19 语料库来微调 GPT-2,以回答关于 COVID-19 的问题。在对 GPT-2 进行微调后,他们又用 BERT 对 GPT-2 的输出进行了修剪:过滤掉不相关的语料,选择与问题语义最相关的句子作为输出。他们的实验表明,GPT-2 和 BERT 的融合模型在自动回答 COVID-19 问题方面表现最好。在面对中文限定领域的智能问答时,Wang

等^[47]提出了一种基于 BERT 获得问题向量表示的模型,将用户输入的问题与之前的问题进行匹配,然后通过数据库返回答案。实验结果证明,该模型在中文限定领域表现出了良好的性能。

Zhou 等^[48]在 BERT 中注入了一个大型的生物医学知识图,使 BERT 能够学习生物医学问答的相关知识。该知识图由多个子图组成,他们使用子图适配器将特定的知识注入到 BERT 中,对 BERT 进行微调。该方法在多个数据集上取得了最先进的性能。

Chau 等^[49]将 BERT 作为法律 QAS 的句对分类器。具体来讲,他们在越南法律数据集上微调了 BERT,将微调后的模型称为 VNLawBERT,用于越南法律领域的问答。VN-LawBERT 与用越南法律文件数据集的特定知识训练的 BERT 相比,在精度、召回率和 F1 值方面均有显著提升。为了快速查询到相似案例,Zhu 等^[50]将顺时针方法与辅助学习相结合对 BERT 进行微调,使用 BM25 排名函数从中文法律案例库中检索出前 n 个候选案件后,用微调后的 BERT 对检索出的案件进行准确排序。经过大量实验,他们证明了该方法优于一些先进的基线方法。

Zhu 等^[51]使用 RoBERTa 和 BERT 在物联网领域开发了一个问答系统。他们在物联网语料库上对 RoBERTa 和 BERT 进行了无监督训练,以学习更多的物联网领域的知识,提高物联网领域问答的质量。他们的实验表明,预训练的模型可以更好地学习特定的知识,并在特定领域的语料库上训练后提高了特定领域问答的质量。

5.2 开放域问答

在 BERT 最初被用于自动问答系统问答时,对应相同答案的段落被视为独立的训练实例,但比较每段答案的分数是很难实现的。为了解决这个问题,Wang 等^[52]提出了一个多段落的 BERT 模型,对同一问题的所有段落的答案分数进行全局标准化,并利用段落排名器来选择高质量的段落。因此,该模型可以通过全局比较得到更准确的答案。Shonibare^[53]开发了一个基于 BERT 的问答系统,称为 ASBERT。BERT 的自注意层在对特定的词进行编码时可以考虑相邻的词,并对输入进行标记嵌入、位置嵌入和分段嵌入的计算,从而获得输出。对于一个给定的问题句和答案候选句,它们的相似度是由 BERT 和距离函数的联合处理决定的。他的实验表明,ASBERT 优于所用的基线模型。

Yi^[54]提出了一个中文开放域问答模型,该系统主要由 3 个部分组成:实体识别、属性抽取和答案检索。首先,她对 BERT 的 12 层 Transformer 生成的向量进行了动态融合,改进后的 BERT 模型被称为 BERDAT。然后,该系统将 BERDAT 与 BiLSTM-CRF 模型融合进行实体识别。接下来,结合 BERDAT 和 softmax 的分类模型进行属性抽取,最后根据前两步的结果进行答案检索。该模型在 NLPCC-KBQA 数据集上的 F1 分数为 97.54%,与以往最先进的方法相比有显著提升。

在现实生活中,不仅文本问答很重要,语音、图像和视频问答也很重要。为了处理关于图像和视频的问答,Yang 等^[55]提出了一种基于 BERT 的视频问答系统。该系统利用

BERT 提取视频中字幕的语义信息,根据提取的字幕语义和视觉特征共同预测答案。他们的实验表明,在 TVQA 数据集上,加入 BERT 后,他们的模型准确性提高了 4.23%。Li 等^[56]利用 XLNet 的交叉编码器,将文本中检测到的实体与社区问答系统的知识库中的相应实体联系起来。他们的实验表明,与一些先进的实体链接(Entity Linking, EL)模型相比,他们的模型具有最高的链接精度。

6 文本摘要

本节将讨论基于 PLM 的文本摘要(Text Summarization, TS)方法。文本摘要指从一个文档中提取关键信息,从而概括出这个文档的主要内容。

6.1 医学领域的文本摘要

自 COVID-19 突然爆发以来,与该病毒相关的文献数量迅速增加,研究人员的阅读速度无法跟上相关文献的增长速度,无法及时掌握 COVID-19 的最新进展。为了缓解这一问题,Kieuvongngam 等^[57]提出了一个基于 BERT 和 GPT-2 的文本摘要模型。他们使用 BERT 将文档中的每个句子转换成 768 维的向量。然后,他们用 K-中心点聚类算法对这些向量进行聚类,提取与关键词配对的摘要,并将其输入 GPT-2 进行训练以生成摘要结果。他们的实验表明,通过整合 BERT 和 GPT-2,他们的模型可以根据关键词生成更精确的摘要。

6.2 法律领域的文本摘要

在法律领域,普通人和专业人士都不可避免地要查阅以前的相似案例。为了帮助普通人和律师获取案件判决书中的关键信息,Liu 等^[58]提出了一种总结法律案件文件的方法。首先,他们提取了案件判决书的 5 个关键部分,将文本摘要问题变成了 5 个不同类别的文本压缩和整合问题。然后,他们为 5 个关键部分微调了 GPT-2,得到了 5 个模型。接下来,他们使用 5 个微调的 GPT-2 模型来对 5 个部分的文本进行总结。最后,他们把 5 个部分的总结放在一起,得到整个案件判决书的完整摘要。他们做了大量的实验,证明了其方法的有效性。

Yoon 等^[59]提出了一个基于 BERT 和 GPT 的韩国法律判决书的摘要模型。他们使用 BERT 通过双向编码器来预测嘈杂的标记,并使用 GPT 来预测下一个标记。他们的实验表明,就标准 ROUGE-1, ROUGE-2 和 ROUGE-L 而言,他们的模型比 BERT2BERT (BERT 与 seq2seq 模型集成)更有效。

Zhou 等^[60]在 seq2seq 模型的基础上融合了 BERT 和注意力(Attention)机制,构建了一个关于法律裁判文书的摘要模型 BASR。在 ROUGE-1, ROUGE-2 和 ROUGE-L 指标上,seq2seq + Attention 模型比其他基线模型分别平均提高了 7.98%, 5.12% 和 8.79%,而 BASR 又比 seq2seq + Attention 模型分别提高了 6.77%, 4.33% 和 6.34%。但值得注意的是,该模型的泛化能力有待加强。

6.3 波斯语中的文本摘要

Farahani 等^[61]提出了两种波斯语的文本摘要方法,并为波斯语 TS 引入了一个名为 pn-summary 的新数据集。他们

采用的 PLM 是 ParsBERT 模型(在波斯语数据集上训练的 BERT 的家族成员)的编码器-解码器版本。这些模型在 pn-summary 数据集上进行了微调。他们的实验显示了其方法的实用性。由于他们的模型是同类中的第一个,因此该模型可以作为未来工作的基线模型。

7 机器翻译

本节将评论一些基于 PLM 的机器翻译方法。机器翻译被用来将文本从一种语言转换到另一种语言。过去,机器翻译主要依靠各种统计模型,而近几年主要依靠各种神经网络,称为神经机器翻译(Neural Machine Translation, NMT)^[62-63]。神经机器翻译通过一个由编码器和解码器组成的单一神经网络将源句子转换为目标句子^[64]。虽然神经机器翻译弥补了传统机器翻译的许多缺点,但它仍然存在许多问题。例如,当对大规模数据集进行训练时,其计算消耗是巨大的,并且不能非常有效地处理单词。因此,研究人员尝试使用 PLM 来处理这些问题。

7.1 基于单语言预训练模型的机器翻译

为了将 PLM 与神经机器翻译相结合,Weng 等^[65]提出了基于 BERT 和 GPT 的神经机器翻译框架 APT。该框架首先使用适配器将一般知识转换为特定任务的表示,同时预训练 BERT 和 GPT,并将这种表示动态地融合到神经机器翻译中。其次,他们使用知识提取范式将 PLM 学到的方面级和句子级知识持续提取到神经机器翻译中。他们在英译德和中译英的实验中表明,APT 优于基于 Transformer 的相应模型。

Zhang 等^[66]提出了一种基于 BERT 的机器翻译方法,称为 BERT-JAM。首先,BERT-JAM 将 BERT 的多层表示融合到 NMT 模型可以使用的整体表示中。其次,BERT-JAM 可以动态地将 BERT 表示与编码器/解码器表示相结合。最后,他们使用一个三阶段的优化策略对 BERT-JAM 进行微调。该策略可以逐渐消减不同的组件,以便在微调期解决灾难性的遗忘问题。他们的实验结果表明,BERT-JAM 在多个翻译任务上的性能是最好的。

为了确保神经机器翻译训练的稳定性 and 可靠性,Shavarani 等^[67]通过 BERT 获得的密集向量形式的方面级语义信

息来增强神经机器翻译。他们的实验表明,这种方法在不增加计算复杂性的情况下提高了翻译质量。

中文表达中存在成语,而对成语的解释与普通词汇的解释不同,这对机器翻译来说是一个挑战。为了解决这一问题,Briskilal 等^[68]提出了一个基于 BERT 和 RoBERTa 的用于成语和文本分类的集合模型架构。他们用 TroFi 数据集对 BERT 和 RoBERTa 进行了微调,并对这两个 PLM 的输出进行了加权平均,以预测最终的类别。与单一的 PLM 相比,他们的混合模型在测试集上的 F1 值和准确性都提高了 2%。

7.2 基于多语言预训练模型的机器翻译

Chen 等^[69]提出了一个名为 Sixt 的跨语言 NMT 模型,它使用 XLM-R 来初始化编码器和解码器的嵌入。他们的实验表明,Sixt 在 15 个任意语言到英语的翻译测试中优于 CRISS 和 m2m-100(两个强大的多语言 NMT 基线)。Wang 等^[70]利用 XLM-R 来提高神经机器翻译的质量,他们进行了 3 种尝试:1)用 XLM-R 对源语言信息进行编码;2)在目标语言端将引入了源语言信息的 XLM-R 作为解码器;3)在源语言端和目标语言端同时引入 XLM-R。他们的实验结果表明,对于资源丰富的语言类型,在源语言端引入 XLM-R 可以更好地对源语言进行编码,提高翻译质量。而对于低资源语言,在目标语言端或者两端同步引入 XLM-R 能够得到更高质量的翻译结果。

Üstün 等^[71]用训练好的特定语言的噪声适配器增强了 mBERT 的交叉注意力。在训练过程中,允许适配器扩展未经训练的语言。该方法在无监督神经机器翻译上的表现优于最先进的基线模型,在有监督神经机器翻译上也有一定优势。Ma 等^[72]提出了一种新型的多语言 PLM,称为 DeltaLM。它使用现有的 PLM(如 XLM)作为编码器,并增加一个解码器作为编码器的附加层。在对该 PLM 进行微调后,他们直接将其用于神经机器翻译。他们的实验表明,DeltaLM 的性能优于基于其他多语言 PLM 的 NMT 模型。

8 挑战

由于 PLM 的性能非常强大,研究人员已经成功地将其应用于 NLP 的各种下游任务,如表 2 所列。

表 2 PLM 在 NLP 领域的实际应用

Table 2 Practical applications of PLM in the field of NLP

任务	单语言 PLM			多语言 PLM		
	BERT	XLNet	GPT	mBERT	XLM	XLM-R
NER	文献[25-26,29-30]			文献[27]		文献[28]
RE	文献[33-35]					
SA	文献[38,42-44]			文献[40-41]		文献[41-42]
QA	文献[48-53]	文献[56]	文献[46]			
TS	文献[57,59-60]		文献[57-59]			
MT	文献[65-68]		文献[65]	文献[70]	文献[71-72]	文献[69]

然而,将 PLM 付诸实践仍有许多挑战,这些挑战的一些例子如下。

(1)当把 PLM 应用于不同的下游任务时,微调过程是不同的。那么,对于一项具体的任务,应该采取什么样的微调策略?应该对哪些参数进行微调以达到最佳效率?对于一项特定的工作,训练数据集的最佳规模是多少?

(2)在特定领域使用 PLM 时,由于特定领域数据集的特殊性,PLM 可能会学习并放大训练数据集中偏见的影响。

(3)与某一特定主题相关的文献,如 COVID-19,正像滚雪球一样不断增加,那么是否有可能在 PLM 的基础上开发自动文献综述系统呢?自动文献综述系统应该比文本摘要复杂得多,因为文本摘要只是总结了一篇文献,而自动文献综述

系统需要总结众多的论文,对它们进行分类比较,并确定未来的研究方向。因此,自动文献摘要可能不仅需要基于 PLM 的文本摘要方法,还需要基于 PLM 的信息抽取方法。此外,论文的长度也是对 PLM 的一个挑战。与 PLM 相关的文献显示 BERT 是最常用的 PLM。然而,BERT 也有一些实际的缺陷。最关键的是,BERT 能处理的最大文本长度是 512 个字符。虽然可以通过截取文本片段和其他方法来处理长文本,但在处理短文本和长文本之间仍然存在显著的性能差距。

(4)对于信息提取中的命名体识别任务,虽然 PLM 在大规模的语料库中进行了预训练,但是命名体本身有太多的不确定性。首先命名体不是一成不变的,它是不断更新的;其次有些命名体的类别比较模糊,并不是确定的属于唯一类别;最后命名体存在嵌套现象。这些都给 PLM 进行命名体识别造成了困难。

(5)随着语音识别和图像识别的发展,人们经常使用音频、照片和视频来表达自己的情感和观点。同时,不同的语言有不同的发音,即使是同一种语言,不同的人的发音也不尽相同。PLM 在同时处理音频、图像、文本和视频中的信息时面临着重大挑战。因此,用 PLM 从图像或视频中个人的身体表情预测情绪也是一个值得深入探索的方向。

(6)问答系统是一种特殊的人机对话系统,因此在未来可将这些基于 PLM 的方法扩展到对话系统。我们认为最有价值的工作是开发基于 PLM 的电子销售人员对话系统,同时基于 PLM 的网上商店客服对话系统也值得研究。这两种系统可能需要与基于 PLM 的情感分析方法相结合,因为任何人机系统都必须考虑人类用户情感对对话结果的影响。

(7)文本摘要可分为抽取式文本摘要和生成式文本摘要。GPT-2 等已经具有很好的文本生成能力,但我们期望在文本摘要方面可以更合理地融合抽取式和生成式,以更好地模仿人撰写摘要的模式。

(8)机器翻译可以采用多语言 PLM,如 mBERT 和 XLM-R。它们在 100 多种语言中进行预训练,但全球范围内有超过 7100 种语言。对许多低资源类别的语言进行注释和训练需要大量的劳动力,而且成本很高,目前很难实现。因此,现有的跨语言 PLM 还不能满足机器翻译的需求。

结束语 预训练语言模型是处理涉及自然语言理解领域的各种任务的有力工具。为了帮助研究人员快速掌握预训练语言模型的应用现状,本文综述了预训练语言模型在命名实体识别、关系提取、情感分析、自动问答、文本摘要和机器翻译中的各种应用。此外,我们还讨论了预训练语言模型的应用所面临的一些挑战。特别是在未来,使用预训练的语言模型来开发自动文献摘要系统、电子销售员系统和在线客服系统是值得深入研究的方向。我们期望预训练模型能够得到更进一步的发展,并能够更好地应用于更多的自然语言处理任务。

参考文献

[1] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: A survey [J]. Science China Technological Sciences, 2020, 63(10): 1872-1897.

[2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is

all you need [C]// Advances in Neural Information Processing Systems. 2017.

- [3] SHERSTINSKY A. Fundamentals of recurrent neural network (RNN) and long short-term memory(LSTM) network [J/OL]. Physica D: Nonlinear Phenomena, 2020, 404: 132306. <https://doi.org/10.1016/j.physd.2019.132306>.
- [4] KENTON, DEVLIN J, CHANG M W, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of NAACL-HLT. 2019: 4171-4186.
- [5] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>.
- [6] ERHAN D, COURVILLE A, BENGIO Y, et al. Why does unsupervised pre-training help deep learning? [C]// Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010: 201-208.
- [7] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: A survey [J]. Science China Technological Sciences, 2020, 63(10): 1872-1897.
- [8] ZAIB M, SHENG Q Z, EMMA ZHANG W. A short survey of pre-trained language models for conversational ai-a new age in NLP [C]// Proceedings of the Australasian Computer Science Week Multiconference. 2020: 1-4.
- [9] LUO X, YIN S, LIN P. A survey of cross-lingual sentiment analysis based on pretrained models [C]// Proceedings of the 21st International Conference on Electronic Business. 2021: 23-33.
- [10] KALYAN K S, RAJASEKHARAN A, SANGEETHA S. AMMU: A survey of transformer-based biomedical pretrained language models [J]. Journal of Biomedical Informatics, 2022, 2(126): 103982.
- [11] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations [C]// International Conference on Learning Representations. 2019.
- [12] CLARK K, LUONG M T, LE Q V, et al. ELECTRA: Pre-training text encoders as discriminators rather than generators [J]. arXiv: 2003. 10555, 2020.
- [13] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized autoregressive pretraining for language understanding [C]// Advances in Neural Information Processing Systems. 2019.
- [14] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>.
- [15] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners [J]. OpenAI Blog, 2019, 1(8): 9.
- [16] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners [C]// Advances in Neural Information Processing Systems 33(NeurIPS 2020). 2020.
- [17] ZHANG Z, HAN X, LIU Z, et al. ERNIE: Enhanced Language Representation with Informative Entities [C]// Proceedings of the 57th Annual Meeting of the Association for Computational

- Linguistics, 2019;1441-1451.
- [18] JIAO X, YIN Y, SHANG L, et al. TinyBERT: Distilling BERT for natural language understanding [C]// Findings of the Association for Computational Linguistics: EMNLP 2020. 2020; 4163-4174.
- [19] CUI Y, CHE W, LIU T, et al. Pre-training with whole word masking for Chinese BERT [J]. Transactions on Audio, Speech, and Language Processing, 2021, 29; 3504-3514.
- [20] CONNEAU A, LAMPLE G. Cross-lingual language model pre-training [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019; 7059-7069.
- [21] CONNEAU A, KHANDELWAL K, GOYAL N, et al. Unsupervised cross-lingual representation learning at scale [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020; 8440-8451.
- [22] LI L H, YATSKAR M, YIN D, et al. VisualBERT: A simple and performant baseline for vision and language [J]. arXiv: 1908.03557.
- [23] HOU Y T, ABULIZI A, ABUDUKELIMU H. Advances in Chinese Pre-training Models [J]. Computer Science, 2022, 49(7): 148-163.
- [24] GRISHMAN R. Information extraction [J]. IEEE Intelligent Systems, 2015, 30(5): 8-15.
- [25] CAI Q. Research on Chinese naming recognition model based on BERT embedding [C]// 2019 IEEE 10th International Conference on Software Engineering and Service Science. 2019; 1-4.
- [26] CHEN D, SONG H Z, ZHANG J, et al. Entity recognition fusing BERT and memory networks [J]. Computer Science, 2021, 48(10): 91-97.
- [27] ZHANG X, LUO X. A Machine-reading-comprehension method for named entity recognition in legal documents [C]// The 29th International Conference on Neural Information Processing. 2022.
- [28] MENG F, YANG S, WANG J, et al. Creating knowledge graph of electric power equipment faults based on BERT-BiLSTM-CRF model [J]. Journal of Electrical Engineering & Technology, 2022, 17; 2507-2516.
- [29] MOON T, AWASTHY P, NI J, et al. Towards lingua franca named entity recognition with BERT [J]. arXiv: 1912.01389, 2019.
- [30] CHEN S, PEI Y, KE Z, et al. Low-resource named entity recognition via the pre-training model [J]. Symmetry, 2021, 13(5): 786.
- [31] BOUDJELLAL N, ZHANG H, KHAN A, et al. ABioNER: A BERT-based model for Arabic biomedical named-entity recognition [J/OL]. Complexity. <https://doi.org/10.1155/2021/6633213>.
- [32] AGRAWAL A, TRIPATHI S, VARDHAN M, et al. BERT-based transfer-learning approach for nested named-entity recognition using joint labeling [J]. Applied Sciences, 2022, 12(3): 976.
- [33] HAN X, WANG L. A novel document-level relation extraction method based on BERT and entity information [J]. IEEE Access, 2020, 8; 96912-96919.
- [34] QIAO B, ZOU Z, HUANG Y, et al. A joint model for entity and relation extraction based on BERT [J]. Neural Computing and Applications, 2022, 34(5): 3471-3481.
- [35] XU S, SUN S, ZHANG Z, et al. BERT gated multi-window attention network for relation extraction [J]. Neurocomputing, 2022, 492; 516-529.
- [36] BAKSHI R K, KAUR N, KAUR R, et al. Opinion mining and sentiment analysis [C]// 2016 3rd International Conference on Computing for Sustainable Global Development. 2016; 452-455.
- [37] MORAES R, VALIATI J F, NETO W P G O. Document-level sentiment classification: An empirical comparison between SVM and ANN [J]. Expert Systems with Applications, 2013, 40(2): 621-633.
- [38] KONG J, WANG J, ZHANG X. Hierarchical BERT with an adaptive fine-tuning strategy for document classification [J/OL]. Knowledge-Based Systems, 2022, 238; 107872. <https://doi.org/10.1016/j.knosys.2021.107872>.
- [39] LIAO W, ZENG B, YIN X, et al. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa [J]. Applied Intelligence, 2021, 51(6): 3522-3533.
- [40] CHEN S, KONG B. cs@ DravidianLangTech-EACL2021: Offensive language identification based on multilingual BERT model [C]// Proceedings of the 1st Workshop on Speech and Language Technologies for Dravidian Languages. 2021; 230-235.
- [41] JAYANTHY S M, GUPTA A. Sj_aj@ dravidianlangtech-eacl2021: Task-adaptive pre-training of multilingual BERT models for offensive language identification [C]// Proceedings of the 4th First Workshop on Speech and Language Technologies. 2021; 307-312.
- [42] XIA M, ZHENG G, MUKHERJEE S, et al. MetaXL: Meta representation transformation for low-resource cross-lingual learning [C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2021; 499-511.
- [43] ARACI D. FinBERT: Financial sentiment analysis with pre-trained language models [J]. arXiv: 1908.10063, 2019.
- [44] CAO Z, ZHOU Y, YANG A, et al. Deep transfer learning mechanism for fine-grained cross-domain sentiment classification [J]. Connection Science, 2021, 33(4): 911-928.
- [45] SHARMA Y, GUPTA S. Deep learning approaches for question answering system [J]. Procedia Computer Science, 2018, 132; 785-794.
- [46] ONIANI D, WANG Y. A qualitative evaluation of language models on automatic question-answering for COVID-19 [C]// Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. 2020; 1-9.
- [47] WANG C, LUO X. A Legal question answering system based on BERT [C]// 2021 5th International Conference on Computer Science and Artificial Intelligence. 2021; 278-283.
- [48] ZHOU S, ZHANG Y. DATLMedQA: a data augmentation and transfer learning based solution for medical question answering [J]. Applied Sciences, 2021, 11(23): 11251.

- [49] CHAU C N, NGUYEN T S, NGUYEN L M. VNLawBERT: A Vietnamese legal answer selection approach using BERT language model [C]//2020 7th NAFOSTED Conference on Information and Computer Science. 2020;298-301.
- [50] ZHU J, LUO X, WU J. A BERT-Based Two-Stage Ranking Method for Legal Case Retrieval [C]//International Conference on Knowledge Science, Engineering and Management. Cham: Springer, 2022;534-546.
- [51] ZHU H, TIWARI P, GHONEIM A, et al. A collaborative AI-enabled pretrained language model for AIoT domain question answering [J]. Transactions on Industrial Informatics, 2021, 18(5):3387-3396.
- [52] WANG Z, NG P, MA X, et al. Multi-passage BERT: A globally normalized BERT model for open-domain question answering [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language. 2019;5878-5882.
- [53] SHONIBARE O. ASBERT: Siamese and triplet network embedding for open question answering [J]. arXiv:2104.08558, 2021.
- [54] YI S W. Research on knowledge graph question answering based on improved BERT [J]. Computer Science and Applications, 2020, 10(12):2361-2370.
- [55] YANG Z, GARCIA N, CHU C, et al. A comparative study of language transformers for video question answering [J]. Neurocomputing, 2021, 445:121-133.
- [56] LI Y, SHEN W, GAO J, et al. Community question answering entity linking via leveraging auxiliary data [J]. arXiv:2205.11917, 2022.
- [57] KIEUVONGNGAM V, TAN B, NIU Y. Automatic text summarization of COVID-19 medical research articles using BERT and GPT-2 [J]. arXiv:2006.01997, 2020.
- [58] LIU J, WU J, LUO X. Chinese Judicial summarising based on short sentence extraction and GPT-2 [C]//Knowledge Science, Engineering and Management (KSEM 2021). Cham: Springer, 2021;376-393.
- [59] YOON J, JUNAID M, ALI S, et al. Abstractive summarization of Korean legal cases using pre-trained language models [C]//2022 16th International Conference on Ubiquitous Information Management and Communication. 2022;1-7.
- [60] ZHOU W, WANG Z, WEI B. Abstractive automatic summarizing model for legal judgment documents [J]. Computer Science, 2021, 48(12):331-336.
- [61] FARAHANI M, GHARACHORLOO M, MANTHOURI M. Leveraging ParsBERT and pretrained mT5 for Persian abstractive text summarization [C]//2021 26th International Computer Conference. Computer Society of Iran, 2017;1-6.
- [62] BAHDANAU D, CHU K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]//Proceedings of the 3rd International Conference on Learning Representations. 2015;1-15.
- [63] WU Y, SCHUSTER M, CHEN Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation [J]. arXiv:1609.08144, 2016.
- [64] DABRE R, CHU C, KUNCHUKUTTAN A. A survey of multilingual neural machine translation [J]. ACM Computing Surveys (CSUR), 2020, 53(5):1-38.
- [65] WENG R, YU H, HUANG S, et al. Acquiring knowledge from pre-trained model to neural machine translation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020;9266-9273.
- [66] ZHANG Z, WU S, JIANG D, et al. BERT-JAM: Maximizing the utilization of BERT for neural machine translation [J]. Neurocomputing, 2021, 460:84-94.
- [67] SHAVARANI H S, SARKAR A. Better neural machine translation by extracting linguistic information from BERT [C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021;2772-2783.
- [68] BRISKILAL J, SUBALALITHA C N. An ensemble model for classifying idioms and literal texts using BERT and RoBERTa [J]. Information Processing & Management, 2022, 59(1):102756.
- [69] CHEN G, MA S, CHEN Y, et al. Zero-shot cross-lingual transfer of neural machine translation with multilingual pretrained encoders [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021;15-26.
- [70] WANG Q, LI M X, WU S X, et al. A neural machine translation approach based on the cross-language pre-trained language model XLM-R [J]. Journal of Peking University: Natural Sciences Edition, 2022, 58(1):29-36.
- [71] ÜSTÜN A, BERARD A, BESACIER L, et al. Multilingual unsupervised neural machine translation with denoising adapters [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021;6650-6662.
- [72] MA S, DONG L, HUANG S, et al. DeltaLM: Encoder-decoder pre-training for language generation and translation by augmenting pretrained multilingual encoders [J]. arXiv:2106.13736, 2021.



SUN Kaili, born in 1997, postgraduate. Her main research interests include artificial intelligence and sentiment analysis.



LUO Xudong, born in 1963, Ph.D, distinguished professor, Ph.D supervisor. His main research interests include natural language processing, intelligent decision-making, game theory, automated negotiation and fuzzy logic.