

## 基于多粒度对比学习的聊天对话摘要模型

康梦瑶, 刘扬, 黄俊恒, 王佰玲, 刘树龙

引用本文

康梦瑶, 刘扬, 黄俊恒, 王佰玲, 刘树龙. [基于多粒度对比学习的聊天对话摘要模型](#)[J]. 计算机科学, 2023, 50(11): 192-200.

KANG Mengyao, LIU Yang, HUANG Junheng, WANG Bailing, LIU Shulong. [Chat Dialogue Summary Model Based on Multi-granularity Contrastive Learning](#) [J]. Computer Science, 2023, 50(11): 192-200.

---

## 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)

### [骨架数据增强和双重最近邻检索自监督动作识别](#)

Self-supervised Action Recognition Based on Skeleton Data Augmentation and Double Nearest Neighbor Retrieval

计算机科学, 2023, 50(11): 97-106. <https://doi.org/10.11896/jsjcx.230500158>

### [融合无监督SimCSE的短文本聚类研究](#)

Study on Short Text Clustering with Unsupervised SimCSE

计算机科学, 2023, 50(11): 71-76. <https://doi.org/10.11896/jsjcx.220900214>

### [基于对比学习的多关系属性图聚类方法](#)

Clustering Method Based on Contrastive Learning for Multi-relation Attribute Graph

计算机科学, 2023, 50(11): 62-70. <https://doi.org/10.11896/jsjcx.220900166>

### [基于边推断增强对比学习的社交媒体谣言检测模型](#)

Rumor Detection Model on Social Media Based on Contrastive Learning with Edge-inferenceAugmentation

计算机科学, 2023, 50(11): 49-54. <https://doi.org/10.11896/jsjcx.221000043>

### [基于侧信道特征的IPSec VPN闭合性检测方法](#)

IPSec VPN Closure Detection Method Based on Side-channel Features

计算机科学, 2023, 50(10): 308-314. <https://doi.org/10.11896/jsjcx.230500141>

# 基于多粒度对比学习的聊天对话摘要模型

康梦瑶<sup>1,2</sup> 刘扬<sup>1,2</sup> 黄俊恒<sup>1,2</sup> 王佰玲<sup>1,2</sup> 刘树龙<sup>1</sup>

1 哈尔滨工业大学(威海)计算机科学与技术学院 山东 威海 264209

2 哈尔滨工业大学(威海)网络空间安全研究院 山东 威海 264209

(kangmengyao99@163.com)

**摘要** 社交网络的发展在给人们带来便捷的同时也产生了海量的聊天数据,如何从聊天对话中筛选出关键信息成为一大难题。聊天摘要为解决此类问题的有效工具,既不必重复浏览冗长的聊天记录,又可以快速获取重要内容。目前,预训练模型被广泛应用于各种类型的文本,包括非结构化、半结构化和结构化文本。然而,针对聊天对话文本的应用,常见的预训练模型难以捕捉到其独特的结构特征,仍需进一步探索与改进。对此,提出了一种基于对比学习的聊天摘要算法 MGCSum。该算法无需人工标注数据集,便于学习和迁移。首先使用文档频数、词项频数和信息熵构造了针对聊天文本的停用词列表,去除聊天中的干扰信息;其次,从词语和主题两个粒度进行自监督对比学习,识别对话中的结构信息,挖掘聊天中的关键词和不同主题信息。在聊天摘要公开数据集 SAMSum 和金融欺诈对话数据集 FINSum 上进行实验,结果表明,与当前主流的聊天摘要方法相比,该算法在摘要的连贯性、信息量和 ROUGE 评价指标上均有显著提升。

**关键词:** 聊天摘要;对比学习;预训练模型;关键词检测;主题分割

中图法分类号 TP391

## Chat Dialogue Summary Model Based on Multi-granularity Contrastive Learning

KANG Mengyao<sup>1,2</sup>, LIU Yang<sup>1,2</sup>, HUANG Junheng<sup>1,2</sup>, WANG Bailing<sup>1,2</sup> and LIU Shulong<sup>1</sup>

1 School of Computer Science and Technology, Harbin Institute of Technology(Weihai), Weihai, Shandong 264209, China

2 Research Institute of Cyberspace Security, Harbin Institute of Technology(Weihai), Weihai, Shandong 264209, China

**Abstract** While the development of social networks brings convenience, but also generates massive amounts of chat data. How to filter key information from chat conversations has become a major difficulty. Chat summary is an effective tool to solve such problems, as it allows users to quickly obtain important content without having to repeatedly browse through lengthy chat records. Currently, pre-trained models are widely used in various types of text, including unstructured, semi-structured, and structured text. However, for chat dialogue text, common pre-trained models are often unable to capture its unique structural features, and further exploration and improvement are still needed. To address these issues, this paper proposes a chat summary model MGCSum, which based on multi-granularity contrastive learning and does not require manual annotation of the datasets, making it easy to learn and transfer. Firstly, a stop word list for chat text is constructed by using document frequency, term frequency and entropy to remove interference information in chat. Then, self-supervised contrastive learning is performed at the granularity of words and topics to identify the structure of conversation, uncover keywords and distinct topic information in chats. Experimental results on the publicly available chat summary datasets SAMSum and financial fraud dialogue summary dataset FINSum show that, compared to current mainstream chat summary methods, this algorithm significantly improves coherence, information content and ROUGE evaluation metrics.

**Keywords** Chat summary, Contrastive learning, Pre-trained models, Keyword detection, Topic segmentation

## 1 引言

近年来,社交网络与人们的关系越来越密切,人们通过社交网络进行交流、交友、工作、学习和娱乐,尤其是受疫情冲击

后,大量工作、教学、会议等线下活动转移至线上,使得人们更加依赖社交网络工具。据 CNNIC 统计<sup>[1]</sup>,截至 2022 年 12 月,国内网民规模达 10.67 亿,其中使用即时通信的用户比例为 97.2%。随着用户数量的日益增加,社交网络产生的通信

到稿日期:2023-03-31 返修日期:2023-05-09

基金项目:国家重点研发计划(2020YFB2009502);国家自然科学基金(62272129);中央高校基本科研业务费专项资金(HIT.NSRIF.2020098)

This work was supported by the National Key R&D Program of China(2020YFB2009502), National Natural Science Foundation of China(62272129) and Fundamental Research Funds for the Central Universities(HIT.NSRIF.2020098).

通信作者:王佰玲(wbl@hit.edu.cn)

数据也与日俱增,同时,人们被大量的网络信息所淹没。因此,人们迫切希望有一种信息处理工具能对这些信息加以分类,按照兴趣加以筛选,甚至提炼出主要信息形成摘要,以便快速、高效地掌握相关信息。

对话是社交网络中最常见的交流方式之一,整理对话内容形成相关的总结和摘要已成为一项重要工作,主要体现在会议、客户服务、医患对话等场景中。会议是多方合作与分享信息的重要渠道,与会者和非与会者都能通过会议纪要快速掌握会议的主要思想。在线客户服务是企业 and 消费者沟通的桥梁,自动摘要能够浓缩已有的客服对话,快速提供解决方案,同时,销售人员通过提炼销售对话中的重要信息来提高销售水平。医患对话是医疗过程中必不可少的环节,医患对话总结为医务工作者拟定治疗方案打下坚实基础,保证医疗活动的顺利进行。可以看出,对话摘要是对话场景中的重要需求,因此,本文重点关注对话摘要,帮助人们在无须回顾复杂对话背景的情况下,快速掌握对话的核心内容。

聊天是工作与生活中不可或缺的沟通途径,高频的聊天产生了庞大的聊天记录,在查找重要内容时,使用关键字检索的方式会耗费过多的时间和人力成本,效率低下。且聊天记录干扰因素繁杂,存在众多疑问句、省略句等口语化内容,以及实际意义不大的系统提示消息和推送信息,而人工筛查时无法避开这些干扰信息。信息化在高速发展的同时也造成了诸多网络社交隐患,通过社交网络实施犯罪的案件层出不穷,犯罪分子的聊天信息往往湮没在正常的交流信息中,给公安人员提取犯罪事实、打击犯罪活动造成极大的困难。

聊天摘要技术是通过自然语言处理算法来实现自动化、

高效率的聊天记录摘要生成,从而改善人们在日常沟通中处理大量聊天信息的能力与效率;此外,聊天摘要还能保障用户在社交网络中的信息安全,在保护用户隐私的同时摘取聊天的关键内容,方便匿名分享信息。除了个人使用,企业机构或管理部门也需要管理聊天群或聊天记录,通过摘要排除噪声信息,提取关键内容,以便有针对性地进行分析和处理。

聊天摘要属于文本摘要研究的范畴,其特别之处在于摘要的原始文本为对话形式。以往的摘要工作大多基于新闻类或公文类等规范的体裁,文本有一定的书写规则,比如新闻文章遵循一种在新闻业中被称为“倒置金字塔”的写作结构<sup>[2]</sup>,在这种形式中,初始段落包含了最有新闻价值的信息,其次是背景信息和详细信息。然而,在对话场景中产生的对话文本并不适用此类规则,因为对话的关键信息往往是分散的,跨越多条语句,并且依赖于不同的角色,导致对话中信息密度较低。因此,文本摘要的方法并不能直接迁移到对话摘要的研究中。

相比会议和客服对话,网络聊天特别随意,文本往往更短、更非结构化,同时包含许多拼写错误、超链接和缩写<sup>[3]</sup>,聊天主题也随着对话频繁切换,图1给出了一段聊天记录示例,聊天中包含2名角色和3个主题,摘要中的信息分布在不同的角色发言和主题中。为了捕获聊天中的语言结构和主题分布,已有的工作大多使用聚类、聊天分隔、主题建模等方法提取话语的语义<sup>[4]</sup>,并没有考虑更细粒度的关键词挖掘,且大部分方法需要额外的人工注释或者自然语言处理工具来标注数据集,导致模型的实用性不强。

对话关键词分布	对话主题分布
A: 你好, <b>建业集团</b> 7月份的费用票有吗 B: 费用票没有 A: 恩 A: <b>进项票</b> 还差一张 A: 是那个税控盘吗 B: 买 <b>税控盘</b> 的 <b>发票</b> 没有 A: <b>车载信息</b> 那个不就是吗 B: 是啊 A: 你是说 <b>发票</b> 还没开来是吧 B: 对 A: 恩 B: 下午我去 <b>税务大厅</b> 问问,看能不能把 <b>发票</b> 拿回来 A: 恩 B: 麻烦把 <b>公司明细</b> 告诉我,下午我一块拿回来,谢谢	A: 你好, 建业集团7月份的费用票有吗 B: 费用票没有 <b>主题 1</b> A: 恩 A: 进项票还差一张 A: 是那个税控盘吗 B: 买税控盘的发票没有 A: 车载信息那个不就是吗 <b>主题 2</b> B: 是啊 A: 你是说 <b>发票</b> 还没开来是吧 B: 对 A: 恩 B: 下午我去 <b>税务大厅</b> 问问,看能不能把 <b>发票</b> 拿回来 A: 恩 <b>主题 3</b> B: 麻烦把 <b>公司明细</b> 告诉我,下午我一块拿回来,谢谢
对话摘要: 没有 <b>建业集团</b> 买 <b>税控盘</b> 的 <b>发票</b> 、 <b>车载信息</b> 的 <b>发票</b> 没开, <b>B</b> 打算下午去 <b>税务大厅</b> 拿 <b>发票</b> 和 <b>公司明细</b> 。	
<div style="display: flex; justify-content: space-around; width: 100%;"> <span>主题 1</span> <span>主题 2</span> <span>主题 3</span> </div>	

图1 聊天摘要示例

Fig.1 Example of chat summary

对比学习在文本摘要中已经得到了广泛应用。相比传统的监督学习,对比学习不需要大量的标注数据,可以在缺乏大规模标注数据的情况下训练模型,便于模型迁移训练;同时,对比学习要求模型区分正样本和负样本之间的差异,这有助于模型更好地理解聊天文本的特征,从而提高摘要质量。

基于以上研究,本文提出了一种基于多粒度对比学习的聊天对话摘要模型,能够高效地概括社交网络中聊天的中心

思想。其主要贡献如下:

(1)基于聊天文本的随意性,构造特定的停用词列表。使用信息熵确定针对网络聊天的特定停用词列表,以最大程度去除聊天中的噪声,降低摘要的冗余性。

(2)使用对比学习从词语和主题两个层面捕捉聊天关键信息。词语层面通过构造自监督正负例子对检测关键词,主题层面通过引入主题一致性检测目标来划分主题。

(3)提出了一种聊天摘要方法 MGCSum,将对学习融合到序列生成框架中,不需要参考摘要之外的人工注释。

## 2 相关工作

### 2.1 对话摘要

随着自动会议记录的日益丰富,会议摘要首先受到了广泛的关注。Shang 等<sup>[5]</sup>针对会议演讲,提出了一种基于图的无监督生成摘要框架,使用多句压缩技术进行会议摘要。Zheng 等<sup>[6]</sup>提出了基于自适应循环神经网络的分层神经编码器学习语义表示,自适应分割会议对话,最终开发了增强的解码器网络以生成会议的抽象摘要。Zhu 等<sup>[7]</sup>设计了一个角色向量来捕获会议中不同演讲者的差异,并且关注会议的层次结构。2019年,Gliwa 等<sup>[8]</sup>提出了第一个高质量的人工标注的聊天摘要语料库 SAMSum,并进行了各种基线实验,迅速激发了聊天对话摘要这一研究方向,现有的工作基本都在 SAMSum 数据集上进行评测。

语言结构是对话摘要的关键。由于聊天中角色的观点和主题都随着对话转移,因此,如何捕获丰富的语言结构成为对话摘要的一个关键问题。Chen 等<sup>[9]</sup>提出了一种多视图序列到序列模型 Multiview BART,他们对比了结构化视图(主题视图和阶段视图)和一般视图(全局视图和离散视图),如图 2 所示,从主题和阶段两个视图分析一段对话可以发现,结构化视图有助于会话编码器捕获对话的细微差别和丰富的信息。Chen 等<sup>[10]</sup>提出对话中丰富的语言结构是生成对话摘要的关键,于是设计了结构化序列到序列模型,在对话文本中构造语篇关系图和动作图,根据图结构明确对话中的三元组(WHO, DOING, WHAT)话语关系,使用关系图解码器和多粒度解码器结合建模话语以及话语关系,最终组合多层次信息进行摘要。Zou 等提出在聊天对话中话题多种多样,且随着对话的进行会频繁切换,还存在大量省略和疑问句,这些句子高度依赖上下文,由此提出了一种融合上下文信息产生对话摘要的新型无监督神经框架 RankAE,其由一个面向主题的排序策略组成,能够同时根据中心性和多样性选择主题话语。

对话	主体视图	阶段视图
J:嘿!我一直在想你	问候	开始
H:噢,不错		
J:你在干什么?	今天的计划	目的
H:我要睡觉了		
J:我想你,我想见到你	明天的计划	讨论
H:明天要早起上班		
J:那明天呢?		
H:老实说,我明天有点事	周六的计划	
J:好,周六怎么样?		
H:可以,我周六有空	接的时间	
J:我8点来接你?		
H:可以,到时候见。		
摘要: J想H,他们同意J在周六早上8点来接H。		

图 2 SAMSum 的示例对话及其主题视图和阶段视图

Fig. 2 Example of SAMSum with its topic and stage views

此外,利用聊天中的关键词引导摘要生成也受到了研究者的关注。Wu 等<sup>[11]</sup>提出了一个分阶段摘要模型 CODS,在第一阶段自动创建包含用户意图信息和关键短语的摘要草图,作为后续摘要阶段的弱监督信号。Zhao 等<sup>[12]</sup>提出了基于

主题词引导的对话图注意网络(TGDGA),根据主题词将对话建模为一个交互图,采用掩码图自注意机制来整合跨句信息流,增加对话语的关注,从而更好地理解对话。

### 2.2 对比学习

对比学习最早应用于计算机视觉领域,目前,在各种自然语言处理任务中也崭露头角。Fang 等<sup>[13]</sup>提出了基于注意力机制的自监督对比学习编码器,直接在句子而不是 token 水平上对语言模型进行预训练,学习句子级表示,预测两个增广句子是否源于同一句子。Klein 等<sup>[14]</sup>在常识推理中应用自监督对比学习,训练模型学习更深入的单词关系,为常识推理提供更好的泛化属性。Gunel 等<sup>[15]</sup>提出在预训练模型的微调阶段应用对比学习,结合交叉熵损失函数,使得性能在多个小样本学习的数据集上取得显著提升;Gao 等<sup>[16]</sup>提出了一个简单的对比学习框架 SimCSE,对输入文本应用随机 dropout 得到正负样本,结合监督学习,在句子级嵌入表示上取得了一定效果。

在对比学习的研究中,最常用的是三元组损失(Triplet Loss)函数<sup>[17]</sup>,该函数被广泛应用于图像识别任务中。具体细节为构造 3 个样本,其中包含 1 个正样本,2 个负样本。模型的训练目标是将正样本与负样本之间的距离尽可能拉近,同时将负样本与负样本之间的距离尽可能拉远。在文本摘要中,该方法被用于学习如何区分摘要与非摘要的文本。

Xu 等<sup>[18]</sup>提出了一个基于对比学习的抽象摘要模型,其在训练过程中最大化原文、参考摘要与生成摘要之间的相似性,能够保持生成摘要与原文的一致性。Zhong 等<sup>[19]</sup>使用对比学习将摘要生成任务转换为语义文本匹配问题,以句子级任务替代摘要级任务。Wu 等<sup>[20]</sup>设计了一个新的基于 BERT 的同时涵盖语言质量和语义信息量的度量标准,通过无监督对比学习来评估没有参考摘要的摘要质量。Liu 等<sup>[21]</sup>在抽象摘要模型训练中阶段性地引入对比学习,解耦对比学习损失和交叉熵损失。Liu 等<sup>[22]</sup>提出主题感知的对比学习模型 ConDigSum,其将一致性目标和子摘要生成目标作为对话摘要的辅助任务,通过交替学习策略更新参数。Geng 等<sup>[23]</sup>重点关注对话中的角色信息,从 Token 层、回合层和全局层 3 个层级进行角色感知监督对比学习,生成的摘要具备更好的事实一致性。这些方法为对比学习在文本摘要中的应用提供了新的思路 and 方向。

## 3 MGCSum 模型

本文提出的 MGCSum 模型原理如图 3 所示,整个模型分为 3 部分:(1)序列到序列的自回归摘要任务,利用带注意力机制的双向预训练模型处理文本,模型对文本的位置、句法、语意等特征编码,充分理解文本后生成预测摘要;(2)关键词检测,使用对比学习区分每个词语对摘要的贡献,为文本中接近摘要的关键词赋予高权重;(3)主题分割,使用主题一致性检测目标,建模聊天中话语之间的相关性变化,以此获取主题分布情况。通过共享参数,将关键词信息与主题信息融入预训练模型的编码器中,提高生成摘要的信息量和连贯性。

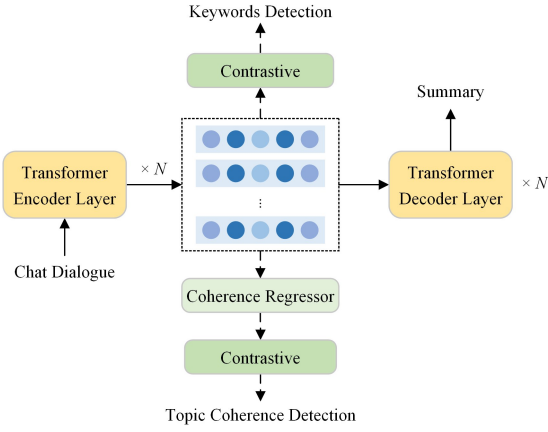


图3 MGCSum 模型

Fig. 3 Model MGCSum

### 3.1 问题定义

给定一个输入的聊天序列  $D = (u_1, u_2, \dots, u_{T_D})$ , 长度为  $T_D$ , 对话摘要的目的是生成  $D$  对应的摘要  $Y_D = (y_1, y_2, \dots, y_{T_y})$ , 摘要长度为  $T_y$ , 需满足  $T_y \ll T_D$ . 为了保留发言者的信息, 将聊天序列中的第  $i$  条聊天表示为  $u_i = (p_i, s_{i1}, s_{i2}, \dots, s_{ij}, \dots)$ , 其中  $p_i$  是第  $i$  条聊天的发言者,  $s_{ij}$  是聊天  $i$  中的第  $j$  句话。

### 3.2 基于 transformer 的序列到序列模型

将聊天对话摘要任务框架看作一个序列到序列的学习问题。我们采用 transformer<sup>[24]</sup> 序列-序列作为主干架构, 模型以对话话语序列作为输入, 并生成相应的摘要。

BART (Bidirectional and Auto-Regressive Transformers) 是基于序列到序列结构的预训练模型<sup>[25]</sup>, 采用了标准的 transformer 整体结构, 并且将激活函数改为 GeLUs 函数。BART 编码器 (Encoder) 为双向特征表示的去噪自编码器, 通过对原始文本进行破坏再重建的方式进行预训练, 计算并提取原文特征; BART 解码器 (Decoder) 为单项特征表示的自回归模型, 其对破坏的文本进行复原重建。二者通过交叉多头注意力连接, 每个解码器层都对编码器最后一层的隐藏状态进行注意力聚合操作, 使得模型的生成文本与原始文本紧密相关。图 4 给出了 BART 预训练模式, 原始文本为 ABCDE, 噪声函数在 B 之前插入一个掩码, 又用掩码替换文本 "CD", 编码器的输入文本变为 "A\_B\_E", 用双向模型进行编码, 解码器结合编码器的输出和原始文本标记, 使用自回归模型来计算原始文本。

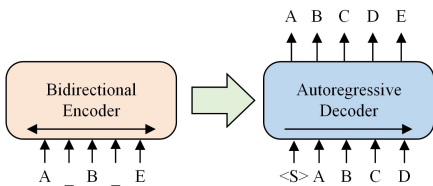


图4 BART 预训练模式

Fig. 4 Pre-training scheme of BART

BART 在摘要生成任务上微调, 在摘要任务中, 输入为聊天序列, 输出为参考摘要。因为 BART 有一个自回归解码器, 所以它可以直接对序列生成任务进行微调, 对于给定的

聊天序列  $D$ , 生成摘要为  $Y_D$ , 训练目标为最小化生成摘要的交叉熵损失, 损失函数定义如式(1)所示:

$$L = - \sum_{t=1}^{T_y} \log p(y_t | y_{<t}, D; \theta) \quad (1)$$

其中,  $y_{<t} = (y_1, y_2, \dots, y_{t-1})$ , 表示输出序列的前  $t-1$  个词;  $\theta$  为训练参数。

BART 将原始对话编码为分布式表示, 解码器基于分布式表示通过注意力机制以及 softmax 技术, 结合 beam search 来生成语意相关且有意义的摘要。具体过程为: 将原始对话经过编码器的处理后, 送入解码器中作为解码阶段的初始状态。首先, 解码器通过自注意力机制, 结合先前生成的部分摘要, 关注输入对话中的重要部分, 确保解码器在生成下一个摘要时, 能够兼顾历史信息, 并同时考虑上下文的影响; 接着, 解码器执行编码器-解码器注意力, 此时, 解码器将自注意力获取的信息与编码器中的信息进行比对, 用于捕捉文本中的关键概念, 作为生成摘要的参考; 然后, 解码器生成下一个词语的概率分布, 使用 softmax 函数进行归一化处理, 从中选择概率最大的词语作为下一个生成的内容; 重复上述过程, 直到生成摘要的长度达到预定的长度或结束标记。为了保证摘要的质量, BART 模型会执行 beam search 操作来搜索最优的输出序列, 解码器在每一步生成预测时, 保留前  $k$  个概率最大的序列作为候选摘要, 最后从中选出一个最优摘要。

### 3.3 领域停用词确定

网络聊天中存在大量的噪声, 这些冗余信息会干扰摘要的生成。在处理聊天文本时, 可以通过删除停用词的方式进行去噪。停用词指在文档中出现频率很高且无实际意义的词, 比如中文词语“的”“也”“啊”等。虽然目前已经有多个训练好的停用词列表, 但是这些现有停用词列表并不完全适用, 因为网络聊天中常出现省略句、缩写词甚至新词。

为了有效识别聊天中的停用词, 去除文本的冗余信息, 本文在现有停用词列表的基础上, 补充针对网络聊天的停用词。综合词语的文档频数、词项频数和信息熵来确定停用词: 文档频数 (Document Frequency, DF) 反映某一词语出现在不同文档中的频率; 词项频数 (Term Frequency, TF) 反映某一词语在给定文档中的出现频率; 信息熵 (Entropy, EN)<sup>[26]</sup> 反映词语在多个文档中出现频率的方差情况, 在某些文档中出现频率非常高但在其他文档中出现频率较低的词语比在每个文档中具有相似频率的词语具有更高的熵。词语  $w$  的信息熵定义如式(2)所示:

$$EN(w) = 1 + \frac{1}{\ln(n)} \sum_{i=1}^n p_i(w) \ln(p_i(w)) \quad (2)$$

其中,

$$p_i(w) = \frac{f_i(w)}{\sum_{j=1}^n f_j(w)} \quad (3)$$

式(2)中,  $n$  表示文档总数; 式(3)中,  $f_i(w)$  表示词语  $w$  在第  $i$  个文档中的词项频数。

DF 和 TF 越高, EN 越低, 词语越有可能成为停用词。综合 3 种方法的合集确定最终的补充停用词列表。

### 3.4 关键词检测

主题分割关注的是对话的粗粒度信息, 而不能区分每个

词语的贡献。因此本文引入了词语级对比学习,在序列到序列模型的编码阶段,使用聊天和摘要构造词语集合,对于每个词语,通过聚类确定它的正例集和负例集,计算正例与负例之间的损失函数,通过反向传播更新模型的权重,最终使模型具备识别对话中特定关键词的能力。

在词语级对比学习中,构造自监督正负例子对。在每次训练开始前,根据给定的输入输出对 $(D_i, Y_{D_i})$ ,从中提取词语集合  $KEY = \{key_{y_1}, key_{y_2}, \dots, key_{y_n}\}$ ,每次训练时,对词语动态地进行聚类,聚类收敛后,对于当前词语 $key_{y_i}$ ,计算该词语与类中每个词语向量的距离,将距离最近的 $N_{key}$ 个词语作为该词语的正例,距离最远的 $N_{key}$ 个词语作为该词语的负例,此处使用余弦相似度来测量两个词嵌入向量之间的距离,定义如式(4)所示:

$$\cos(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2} \quad (4)$$

对于每个关键词 $key_{y_i}$ ,通过 transformer 编码器的最后一层获得它的上下文向量表示 $e_i$ ,对应的某个正例 $key_{y_j}^+$ 和负例 $key_{y_j}^-$ 的向量表示为 $y_j^+$ 和 $y_j^-$ ,此处使用欧氏距离作为距离函数,具体表示如式(5)所示:

$$d(u, v) = \|u - v\|_2 \quad (5)$$

对于当前词语 $key_{y_i}$ ,存在 $N_{key}$ 对正负例,定义损失函数为式(6):

$$L_{key} = \frac{1}{N_{key}} \sum_{j=1}^{N_{key}} \max(0, d(y_i, y_j^+) - d(y_i, y_j^-) + \alpha) \quad (6)$$

其中, $\alpha$ 是边际系数,用于控制正负例之间的距离。当 $d(a_i, pos_i) - d(a_i, neg_i) > 0$ 时,损失函数返回 $d(a_i, pos_i) - d(a_i, neg_i) + \alpha$ ,否则返回0。使用式(7)计算词语 $key_{y_i}$ 的权重 $key_{y_i}^*$ :

$$key_{y_i}^* = \text{softmax}(-L_{key}) \quad (7)$$

得到的词语权重可用于更新编码器中的参数。

### 3.5 主题分割

主题分割是对话摘要中重要的处理步骤,主题分割指将对话中的文本划分为不同的主题或话题,以便更好地理解对话的主题结构和内容,更好地提取关键信息,从而生成更准确、更有用的摘要。常用的主题分割方法通常需要额外的人工注释,为不同的对话主题标注标签,或将主题分割视为一个独立的子任务,导致传播过程中误差累积。本文使用对比学习的方法,在序列到序列模型的编码阶段,使用聊天文本,将原始聊天序列作为正例,打乱顺序的聊天序列作为负例,计算正例与负例之间的损失函数,通过反向传播更新模型的权重,使模型具备捕获话语连贯性信息的能力,从而进行对话的主题分割。

假设同一主题内的话语比不同主题间的话语更连贯,在此基础上引入主题一致性检测目标。根据定义,一个对话包含 $T_D$ 条聊天话语,使用固定大小的窗口,选择对话 $D$ 中连续的 $k$  ( $k < T_D$ )条子序列 $(u_{i+1}, u_{i+2}, \dots, u_{i+k})$ 作为聊天片段的正例 $P_k$ ,对应的负例 $N_k$ 通过随机变换正例中聊天句子的顺序来构造,二者的上下文向量表示为 $E_k^+$ 和 $E_k^-$ ,聊天片段的主题一致性得分如式(8)所示:

$$[\text{con}(P_k), \text{con}(N_k)] = \text{softmax}([\mathbf{y}_k^+, \mathbf{y}_k^-]) \quad (8)$$

其中,

$$\mathbf{y}_k^+ = \omega_1 * \mathbf{E}_k^+ + b_1, \mathbf{y}_k^- = \omega_2 * \mathbf{E}_k^- + b_2 \quad (9)$$

其中, $\omega_1$ 和 $b_1$ 为训练参数。一个对话 $D$ 至少存在 $|D-k|$ 个正负例子对,每次训练时随机选择其中的 $N_s$ 对样本,损失函数如式(10)所示:

$$L_{\text{con}}^D = \frac{1}{N_s} \sum_{j=1}^{N_s} \max(0, \beta - (\text{con}(P_k) - \text{con}(N_k))) \quad (10)$$

其中, $\beta$ 是边际系数,用于控制聊天片段正负例之间的距离; $k, N_s, \beta$ 均为超参数。

## 4 实验

### 4.1 实验数据集

本文在两个数据集上评估 MGCSum 的有效性。SAM-Sum 是由语言学家编写的英文对话摘要数据集,一共包含 16000 个对话的抽象摘要,有 14732 个对话摘要对用于训练,818 个实例用于验证,819 个实例用于测试,训练集中每个对话都由语言学家编写摘要。FINSum 数据集是一个金融欺诈对话数据集,源于某经侦部门提供的社交网络聊天数据,包含 300000 条聊天记录和人工编写的摘要,涉及金融欺诈、讨论事件和日常生活等方面的内容,按照 80%,10%,10% 的比例划分为训练集、验证集和测试集 3 个部分,每个部分的具体统计信息如表 1 所列。

表 1 FINSum 数据集特征

Table 1 Features of dataset FINSum

数据集	对话平均 句子数量	对话句子 数量范围	对话平均 长度	摘要平均 长度
训练集	14.71	[5,36]	150.17	53.85
验证集	16.67	[5,26]	174.35	51.90
测试集	13.88	[5,27]	145.70	47.20

### 4.2 评价指标

本文采用自动评估和人工评估两种评价方式。

自动评估:为了评估本文的模型,使用 ROUGE<sup>[27]</sup>的 $F_1$ 值来衡量由不同模型生成摘要的质量,ROUGE 计算生成摘要和参考摘要之间的 $N$ 元组重叠程度, $F_1$ 的计算式如式(11)所示:

$$F_1 = \frac{2 R_{\text{ROUGE-N}} P_{\text{ROUGE-N}}}{R_{\text{ROUGE-N}} + P_{\text{ROUGE-N}}} \quad (11)$$

其中, $R_{\text{ROUGE-N}}$ 表示召回率, $P_{\text{ROUGE-N}}$ 表示准确率,二者的计算公式如式(12)和式(13)所示。

$$R_{\text{ROUGE-N}} = \frac{\sum_{S \in \{\text{RefSummary}\}} \sum_{gram_n \in S} \text{Count}_{\text{match}}(gram_n)}{\sum_{S \in \{\text{RefSummary}\}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (12)$$

$$P_{\text{ROUGE-N}} = \frac{\sum_{S \in \{\text{RefSummary}\}} \sum_{gram_n \in S} \text{Count}_{\text{match}}(gram_n)}{\sum_{S \in \{\text{GenSummary}\}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (13)$$

实验使用 ROUGE-1, ROUGE-2 和 ROUGE-L 这 3 个评估指标,采用基于 python 的 files2rouge 包<sup>1)</sup>来计算 $F_1$ 值。

人工评估:由于 ROUGE 重点关注生成摘要与参考摘要之间的字面匹配,并不能确定生成的摘要是否事实一致或者合乎语法,因此本文还采用人工评估来衡量摘要质量。从事实一致性、语意连贯性、信息量和简洁性 4 个方面进行打分,

<sup>1)</sup> files2rouge; <https://github.com/pltrdy/files2rouge>

事实一致评估生成摘要与原文的语意一致,语意连贯评估生成摘要语序通顺以及衔接自然程度,信息量评估生成摘要包含关键信息的程度,简洁性评估摘要丢弃冗余信息的程度。

### 4.3 实验设置

本文的实验硬件环境为 Tesla P40(24 GB) 单个 GPU。模型主干架构是序列到序列框架,使用 BART-large 作为预训练模型,编码器和解码器均为 12 层,隐藏层大小为 1024,前馈滤波大小为 4096,BART 每层有 16 个注意力头,模型的参数大小为  $400 \times 10^6$ ,所有层的 Dropout 随机失活率设置为 0.1,采用 Adam 优化器作为优化算法。SAMSum 和 FINSum 的学习率分别为 0.00004 和 0.00005,两个对比目标的边际系数  $\alpha$  和  $\beta$  设置为 1,其他的超参数在训练过程中进行调整,采用交替参数更新策略进行训练。

### 4.4 实验结果及分析

#### 4.4.1 对比模型

为了验证本文提出的 MGCSum 模型的有效性,将其与主流摘要模型进行对比。

Lead3 是摘要任务中最常用的一种简单高效的方法,它直接将文本的前三个主要句子作为摘要。

TextRank 是一种基于图的无监督排名算法,该算法将句子作为节点,将句子之间的相似度作为边,通过句子之间推荐迭代来抽取关键句组成摘要。

Pointer Generator 是基于注意力机制的序列到序列模型,使用指针生成器网络复制原文内容,使用覆盖损失策略避免摘要重复。

BART 是一个预训练的编码-解码 transformer 模型,有 base 和 large 两个版本,对应不同规模的训练数据与模型参数,文中 BART 表示的是 large 版本。

DialoGPT<sup>[28]</sup> 是一个在开放域 Reddit 数据上预先训练过的自回归对话系统,使用多层 transformer 捕获对话中的联合概率分布,生成种类丰富的句子。DynamicConv<sup>[29]</sup> 提出了一种轻量级的动态卷积来取代 transformer 中的自注意模块,避免输入文本过长导致计算量过大的问题。

FastAbs-RL<sup>[30]</sup> 是一个准确快速的抽象摘要模型,其对原文中的重要句子进行压缩和改写,通过强化学习生成抽象摘要。

D-HGN<sup>[31]</sup> 是一个异构对话图网络模型,它引入大规模常识知识库,为对话中的发言者、内容和常识构建异构对话图,该摘要模型具有不错的泛化能力。

UniLM<sup>[32]</sup> 是一个预训练的序列到序列语言模型,能够在摘要生成任务上进行微调。

Multiview BART 是一个多视图序列到序列模型,从对话的主题、阶段、全局和离散角度提取对话结构,使用多视图解码器合并不同的视图来生成对话摘要。

SICK<sup>[33]</sup> 是一个常识性知识模型,它采用外部常识知识模型,在自然语言中生成一组丰富的常识推论,缓解对话上下文信息不足的问题。在 SICK 的基础上,SICK++ 使用从参考摘要中生成的常识作为监督,与对话摘要任务共同学习,将常识添加到编码器中。

CODS 是一个两阶段对话摘要方法,第一阶段识别对话

意图和关键短语,并将其作为第二阶段摘要的弱监督信号,同时控制摘要粒度。

#### 4.4.2 对比实验

本文将 MGCSum 与上述对比模型分别在 SAMSum 和 FINSum 数据集上进行实验,使用 4.2 节中的评价指标,对比结果如表 2 和表 3 所列,每列中最好的结果加粗标注。

表 2 基于 SAMSum 数据集的 ROUGE 结果  
Table 2 ROUGE results on test set SAMSum

模型	ROUGE-1	ROUGE-2	ROUGE-L
Lead3	19.32	10.78	19.12
TextRank	22.16	11.34	21.77
PointerGenerator	32.27	14.42	34.36
DialoGPT	39.77	16.58	38.42
DynamicConv	41.07	17.11	37.27
FastAbs-RL	41.03	16.93	39.05
D-HGN	42.03	18.07	39.56
Multiview BART	49.30	25.60	47.7
UniLM	50.00	26.03	42.34
SICK	53.39	28.42	49.12
SICK++	53.73	28.81	49.50
CODS	52.65	27.84	<b>50.79</b>
MGCSum	<b>53.84</b>	<b>28.49</b>	44.38

表 3 基于 FINSum 数据集的 ROUGE 结果  
Table 3 ROUGE results on test set FINSum

模型	ROUGE-1	ROUGE-2	ROUGE-L
Lead3	24.85	9.61	23.89
TextRank	31.49	10.07	26.23
PointerGenerator	35.38	17.32	29.55
ConDigSum	44.61	18.86	30.30
BART	43.84	17.47	27.79
MGCSum	<b>45.48</b>	<b>19.23</b>	31.30
w/o Keywords	44.99	19.09	<b>32.08</b>
w/o TopicSeg	44.88	19.15	30.95

自动评估:从表 2 和表 3 的 ROUGE 结果中可以看出,使用预训练模型的方法比没有经过预训练的方法表现更好,这证明了预训练模型在对话摘要任务中的有效性。本文提出的模型 MGCSum 在两个数据集上的性能均有提升,在 SAMSum 数据集上 ROUGE-1 和 ROUGE-2 分别提升 1.2 和 0.6,在 FINSum 数据集上表现更好,ROUGE-1 提高 2.8,ROUGE-2 提高 1.4,ROUGE-L 提高 1.0,各项 ROUGE 结果显著优于现有模型。相比 SAMSum 数据集,FINSum 数据集主要涉及金融领域与日常交流,有较多专业相关词汇,文本区分度更高,由于对比学习重点关注文本之间的差异性和相似性,所以模型在 FINSum 数据集上的表现更好,说明 MGCSum 模型在特定领域的对话摘要问题上应用性更强。

本文在 FINSum 数据集上进行了消融实验,结果显示主题分割对于摘要的提升效果更好,这是因为主题分割捕捉到了聊天的结构信息与话题转移,增强了模型对聊天的理解。关键词检测算法在 ROUGE-2 和 ROUGE-L 上的提升表现不如主题分割,本文认为原因是预训练模型中已经包含 Token 级别的权重训练,因此关键词检测算法对摘要模型的提升有限。

人工评估:从数据集 FINSum 的测试集中随机抽取 20 个

聊天摘要,每个指标从1(最差)到5(最好)进行打分。为了减少由人工分歧引起的误差,每个摘要都由3位具备自然语言处理背景的评估者进行评分,结果如表4所列。从实验结果可以看出,本文提出的模型MGCSum在语意连贯性、信息量和简洁性3个指标上都得到了更高的分数,相比其他抽象式摘要模型,事实一致性的分数也更高。抽取式摘要模型(Lead3和TextRank)产生的摘要可以保持事实一致性,但是语意并不连贯,包含的信息量也不足,简洁性取决于抽取的句子数量;生成式摘要模型(PointerGenerator, BART, MGCSum)的事实一致性较差,但语意比较连贯,覆盖的信息量也更大。本文还对消融模型的摘要结果同样进行了人工评估,只加入主题分割算法,能够取得最高的语意连贯性和简洁性分数,但摘要覆盖的信息量不足;只加入关键词检测算法,可以保证摘要信息量充足,但连贯性和简洁性的分数较低,说明关键词检测与主题分割分别起到了不同作用。

表4 基于FINSum数据集的人工评估结果

Table 4 Human evaluation results on test set FINSum

模型	事实一致性	语意连贯性	信息量	简洁性
Gold	4.93	4.75	4.92	4.58
Lead3	4.25	3.01	2.42	3.51
TextRank	<b>4.33</b>	2.75	3.17	2.75
PointerGenerator	3.58	2.83	3.25	2.67
BART	3.67	3.83	3.67	3.33
MGCSum	3.92	4.02	<b>4.25</b>	3.75
w/o Keywords	3.83	<b>4.08</b>	3.33	<b>4.17</b>
w/o TopicSeg	3.85	3.42	4.08	3.25

#### 4.4.3 对比学习分析

##### (1) 关键词检测模块

关键词检测作为本文模型的一个重要模块,能够使模型具备识别关键词的能力。为了验证模块的有效性,本文从数据集FINSum的测试集中随机抽取了40个聊天摘要,使用准确率、精准率、召回率和F1值来评价BART模型与MGCSum模型生成摘要中关键词的识别情况,结果如图5所示。可以看出,添加关键词检测模块后,模型在关键词检测上的各个

指标均得到了提升。

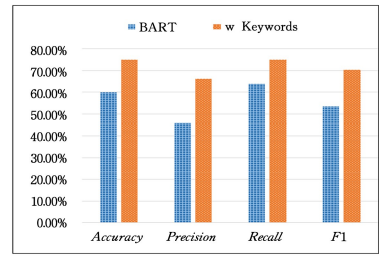


图5 关键词检测实验

Fig. 5 Results of keyword detection

##### (2) 主题分割模块

主题分割是获取聊天文本对话结构的关键步骤,模型通过对比学习正常的聊天序列与随机打乱的聊天序列,获取句子之间的语意连贯性,从而关注聊天中的主题变换情况。为了验证模块的有效性,本文从FINSum的测试集中随机选取3个聊天依次连接,以固定的窗口遍历聊天,检测窗口内聊天文本的主题一致性得分,结果如图6所示,横轴中 $\{S_i, E_i\}$ 表示第*i*段聊天的起始与结束。可以看出,原始聊天片段中一致性得分高,不同聊天片段之间的一致性得分低,说明主题一致性检测目标能够进行句子间语意连贯性检测,以捕捉对话中的主题分布。

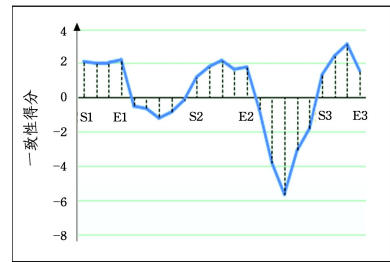


图6 不同聊天对话的一致性得分

Fig. 6 Coherence scores of different chat dialogues

图7给出了一个MGCSum生成的聊天摘要与基线模型BART生成的摘要以及参考摘要的比较。

聊天对话			
A: 王姐, 建业是啥时候来的?	B: 我也不知道呢	主题 1	
B: 我去的时候他就要在那报了	A: 噢噢 我看了看7月份的凭证好像没有		
A: 他不在这报了, 要走了	B: 哦, 那就有给他哪些就行	主题 2	
A: 王姐 建业7月份的凭证系统里有纸制的没找着。	B: 他当时是不是拿走来, 查账的时候		
A: 啥时候查账来?	B: 八九月份的吧	主题 3	
B: 那时候我刚去, 也是不太清楚, 问问赵老师	A: 那他写条来么		
B: 没有吧	B: 赵老师给拿的		
			BART: A和B讨论了建业的情况, B建议给他, 但A说7月系统中的纸质凭证, 但B找不到, B要求A写一张纸条, 但A没有写。 R-1:44.82 R-2: 5.41 R-L:35.01
			MGCSum: (主题 1) A和王姐讨论了建业的情况和(主题 2) 7月份的凭证, A发现没有找到交易系统里的纸质凭证, (主题 3) B也不知道, 说是赵老师拿的, 并且没有写条。 R-1:48.27 R-2:19.18 R-L:36.84
		参考摘要: A询问王姐建业来的时间, A发现系统里有7月份的凭证, 但没有纸质的, 王姐提出可能是八九月份查账时拿走了, 建议A问问赵老师。	

图7 聊天摘要结果示例

Fig. 7 Examples of chat summaries

从图中可以看出,两人的聊天对话按顺序依次包含3个主题,第一个主题是建业的报账时间,第二个主题是系统已有的凭证,第三个主题是纸质凭证的可能去处,BART生成的摘要遗漏了“王姐”“系统”“赵老师”等关键信息,语句不太

连贯,且第三个主题的意思与原文完全不符,发生了一致性错误;MGCSum生成的摘要则识别了主要的关键词以及3个主题信息,这进一步说明本文提出的模型MGCSum能够捕捉对话的主题信息,覆盖更多的核心内容,同时,语意更加连贯

且与原文一致,但与参考摘要仍存在较大差距。

综上所述,在聊天摘要中引入对比学习能够提高模型的泛化能力,对比学习可以通过对不同样本进行比较,发现样本之间的差异和相似性,从而更好地区分并学习不同的特征。此外,自然语言处理任务中普遍存在训练集数据稀少、标注错误、噪声数据等问题,这些问题往往会导致模型的性能表现下降,对比学习能够对文本数据进行有效的特征提取和分类,有助于提升模型的性能表现。

**结束语** 本文在序列到序列模型的基础上,使用对比学习从词语和主题两个粒度进行自监督训练,增强了聊天摘要的信息量与连贯性,并且识别了社交网络聊天中常见的停用词,这种去噪方式可以有效提高对话摘要的质量,减少干扰因素对结果的影响。实验结果表明,MGCSum模型能够有效捕获聊天中的关键信息。后续工作将在更多的聊天数据集上进行测试,验证模型的有效性。从人工评估中可以发现模型存在抽象摘要的通病,生成的摘要不能保持较好的事实一致性,未来可考虑在模型训练过程中引入事实一致性度量指标,或者结合抽取式摘要方法,提高摘要的忠实度。

### 参 考 文 献

[1] 中国互联网络信息中心. 第51次《中国互联网络发展状况统计报告》. [EB/OL]. <https://www.cnnic.cn/n4/2023/0302/c199-10755.html>.

[2] KRYSZCINSKI W, KESKAR N S, MCCANN B, et al. Neural Text Summarization: A Critical Evaluation[C]// Conference on Empirical Methods in Natural Language Processing & International Joint Conference on Natural Language Processing. 2019: 540-551.

[3] KOTO F. A publicly available Indonesian corpora for automatic abstractive and extractive chat summarization[C]// International Conference on Language Resources and Evaluation. 2016: 801-805.

[4] ZOU Y, LIN J, ZHAO L, et al. Unsupervised summarization for chat logs with topic-oriented ranking and context-aware auto-encoders[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(16): 14674-14682.

[5] SHANG G, DING W, ZHANG Z, et al. Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 664-674.

[6] ZHENG J, ZHAO Z, SONG Z, et al. Abstractive meeting summarization by hierarchical adaptive segmental network learning with multiple revising steps [J]. *Neurocomputing*, 2020, 378: 179-188.

[7] ZHU C, XU R, ZENG M, et al. End-to-end abstractive summarization for meetings [J]. *arXiv*:2004.02016, 2020.

[8] GLIWA B, MOCHOL I, BIESEK M, et al. SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization[C]// Workshop on New Frontiers in Summarization. 2019: 70-79.

[9] CHEN J, YANG D. Multi-View Sequence-to-Sequence Models

with Conversational Structure for Abstractive Dialogue Summarization[C]// Conference on Empirical Methods in Natural Language Processing. 2020: 4106-4118.

[10] CHEN J, YANG D. Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs[C]// Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021: 1380-1391.

[11] WU C S, LIU L, LIU W, et al. Controllable abstractive dialogue summarization with sketch supervision[J]. *arXiv*: 2105.14064, 2021.

[12] ZHAO L, XU W, GUO J. Improving abstractive dialogue summarization with graph structures and topic words[C]// International Conference on Computational Linguistics. 2020: 437-449.

[13] FANG H, WANG S, ZHOU M, et al. Cert: Contrastive self-supervised learning for language understanding[J]. *arXiv*: 2005.12766, 2020.

[14] KLEIN T, NABI M. Contrastive Self-Supervised Learning for Commonsense Reasoning[C]// Annual Meeting of the Association for Computational Linguistics. 2020: 7517-7523.

[15] GUNEL B, DU J, CONNEAU A, et al. Supervised contrastive learning for pre-trained language model fine-tuning[J]. *arXiv*: 2011.01403, 2020.

[16] GAO T, YAO X, CHEN D. SimCSE: Simple Contrastive Learning of Sentence Embeddings[C]// Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics. 2021: 6894-6910.

[17] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 815-823.

[18] XU S, ZHANG X, WU Y, et al. Sequence level contrastive learning for text summarization[C]// Conference on Artificial Intelligence. 2022: 11556-11565.

[19] ZHONG M, LIU P, CHEN Y, et al. Extractive Summarization as Text Matching[C]// Annual Meeting of the Association for Computational Linguistics. 2020: 6197-6208.

[20] WU H, MA T, WU L, et al. Unsupervised Reference-Free Summary Quality Evaluation via Contrastive Learning[C]// Conference on Empirical Methods in Natural Language Processing. 2020: 3612-3621.

[21] LIU Y, LIU P. SimCLS: A Simple Framework for Contrastive Learning of Abstractive Summarization[C]// Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021: 1065-1072.

[22] LIU J, ZOU Y, ZHANG H, et al. Topic-Aware Contrastive Learning for Abstractive Dialogue Summarization[C]// Association for Computational Linguistics: EMNLP. 2021: 1229-1243.

[23] GENG Z, ZHONG M, YIN Z, et al. Improving Abstractive Dialogue Summarization with Speaker-Aware Supervised Contrastive Learning[C]// International Conference on Computational Linguistics. 2022: 6540-6546.

- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [25] LEWIS M, LIU Y, GOYAL N, et al. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[J]. *arXiv*: 1910. 13461, 2019.
- [26] SINKA M P, CORNE D W. Towards modernised and web-specific stoplists for web document analysis[C] // *International Conference on Web Intelligence. IEEE*, 2003: 396-402.
- [27] LIN C Y. Rouge: A package for automatic evaluation of summaries[C] // *Proceedings of the Workshop on Text Summarization Branches Out*, 2004: 74-81.
- [28] ZHANG Y, SUN S, GALLEY M, et al. Dialogpt: Large-scale generative pre-training for conversational response generation[J]. *arXiv*: 1911. 00536, 2019.
- [29] WU F, FAN A, BAEVSKI A, et al. Pay less attention with lightweight and dynamic convolutions[J]. *arXiv*: 1901. 10430, 2019.
- [30] CHEN Y C, BANSAL M. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting[C] // *Annual Meeting of the Association for Computational Linguistics*, 2018: 675-686.
- [31] FENG X, FENG X, QIN B. Incorporating commonsense knowledge into abstractive dialogue summarization via heterogeneous graph networks[C] // *Chinese Computational Linguistics*; 20th China National Conference. Cham: Springer International Publishing, 2021: 127-142.
- [32] DONG L, YANG N, WANG W, et al. Unified language model pre-training for natural language understanding and generation[J]. *Advances in Neural Information Processing Systems*, 2019, 32: 13063-13075.
- [33] KIM S, JOO S J, CHAE H, et al. Mind the Gap! Injecting Commonsense Knowledge for Abstractive Dialogue Summarization[J]. *arXiv*: 2209. 00930, 2022.



**KANG Mengyao**, born in 1999, post-graduate. Her main research interests include artificial intelligence, natural language processing and cyber security.



**WANG Bailing**, born in 1978, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include cyber security and industrial Internet security.

(责任编辑:何杨)