计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202220951
61(5): 1299−1309, 2024

融入知识图谱和课程学习的对话情绪识别

杜金明　孙媛媛　林鸿飞　杨　亮
（大连理工大学计算机科学与技术学院　辽宁大连　116024）
（dujinming@mail.dlut.edu.cn）

Conversational Emotion Recognition Incorporating Knowledge Graph and Curriculum Learning

Du Jinming, Sun Yuanyuan, Lin Hongfei, and Yang Liang
（*College of Computer Science and Technology*, *Dalian University of Technology*, *Dalian*, *Liaoning* 116024）

Abstract　Conversational emotion recognition is the task of classifying emotions based on conversations. The conversation data are characterized by colloquial language and a wide range of topics, with semantic similarities among labels. Colloquial language exhibits issues such as word ambiguity and the omission of semantic information, emphasizing the importance of common sense and grammatical knowledge in conversational emotion recognition tasks, and these factors enable the model to accurately capture semantic information. Moreover, the current challenge lies in the variations in text richness and the frequency of emotion transfer across different dialogue scenarios, which result in suboptimal classification performance. We propose CK-ERC model to address these challenges. In the pre-training phase, CK-ERC model extracts structured data to incorporate common sense knowledge graphs and grammatical knowledge graphs, aiding the model in accurately capturing colloquial information. In the fine-tuning phase, a supervised contrast learning task is introduced to help the model identify similar emotional labels. Furthermore, a dynamic threshold-based curriculum learning strategy is designed for training and optimizing the model based on text richness (from high to low) and emotion transfer frequency (from low to high). CK-ERC model demonstrates superior performance in various conversation modes, including two-person conversation, multi-person conversation, simulated conversation, and daily conversation. Particularly, CK-ERC model achieves the best performance on MELD and EmoryNLP datasets.

Key words　conversational emotion recognition; contrast learning; knowledge graph; curriculum learning; transfer learning

摘　要　对话领域情绪识别是基于对话的情感分类任务，对话数据具有口语化、主题跨度大和标签具有语义相似性的特点. 口语化表现为对话中存在隐含常识和语法知识的二义性词语和省略句，导致模型难以准确建模语义信息；主题跨度大表现为不同对话场景下的文本信息丰富度差异大、情绪转移频率差异大，导致模型性能下降. 提出 CK-ERC 模型缓解上述问题，在预训练阶段，抽取结构化数据为模型融入常识和语法知识图谱，帮助模型建模口语化信息；在微调阶段引入监督对比学习任务帮助模型识别相似情绪标签；在训练策略上设计了基于动态阈值的课程学习策略，按照文本丰富度从高到低、情绪转移频率从低到

收稿日期：2022-11-18；**修回日期**：2023-07-24
基金项目：国家重点研发计划项目(2022YFC3301801)；中央高校基本科研业务费专项资金(DUT22ZD205)
This work was supported by the National Key Research and Development Program of China (2022YFC3301801) and the Fundamental Research Funds for the Central Universities (DUT22ZD205)
通信作者：孙媛媛(syuan@dlut.edu.cn)

高的策略优化模型. CK-ERC 模型在双人对话、多人对话、模拟对话、日常对话等多种对话模式下显著优于其他模型，在 MELD 和 EmoryNLP 数据集上获得最佳表现.

关键词 对话情绪识别；对比学习；知识图谱；课程学习；迁移学习

中图法分类号 TP391

随着人工客服、智能语音等对话任务的快速发展，仅仅能完成用户指定任务的对话系统已经不能满足用户的要求，对话系统需要考虑用户的情绪，多项综述文献 [1−2] 介绍了情绪识别任务的重要性和难点. 对话领域情绪识别是基于对话的情感分类任务，更加贴近应用场景，具有很大的研究意义. 同时，对话数据具有口语化、主题跨度大和标签具有语义相似性的特点，为对话情绪识别任务带来很大的挑战.

第 1 个挑战是：如何帮助模型准确建模口语化信息. 对话参与者具有知识储备，导致对话中存在二义性词语和省略句，如图 1 所示，“It’s a piece of cake”是二义性语句，“Believe it or not”是省略句. 因此，为对话情绪识别模型融入外部知识的研究备受关注，尹庆宇等人[3] 通过帮助模型恢复省略句来优化模型性能. Ghosal 等人[4] 和 Zhang 等人[5] 分别为基于递归神经网络和图神经网络的对话情绪识别模型融入常识知识，优化模型性能. 张晓宇等人[6] 为对话模型融入外部知识，帮助模型准确建模对话信息. 文献 [3−6] 证明了融入外部知识能够帮助模型建模语义信息. 随着预训练模型在自然语言处理任务中取得优异表现，为基于预训练模型的对话情绪识别任务融入外部知识成为值得探究的课题. 但为预训练模型融入外部知识存在异构向量难以融合、知识灾难性遗忘等问题，对话情绪识别领域尚未发现相关研究工作.

第 2 个挑战是：如何应对主题跨度大导致模型性能下降的现象. Zhou 等人[7]、Huang 等人[8]、曾维新等人[9] 将课程学习分别应用于机器翻译、关系提取和实体对齐任务上，模型性能都获得了显著的提升，证明了课程学习策略可以有效解决数据噪声和数据学习难度差异大带来的问题. 因此，本文应用课程学习策略解决主题跨度大导致模型性能下降的问题. 但是，课程学习的重排序策略会破坏对话上下文依赖关系，对话情绪识别任务是在对话场景下进行情绪识别的任务，对话场景在文本中的体现就是对话上下文依赖关系. 那么如何设计不破坏对话上下文依赖关系的课程学习策略是应对挑战的关键.

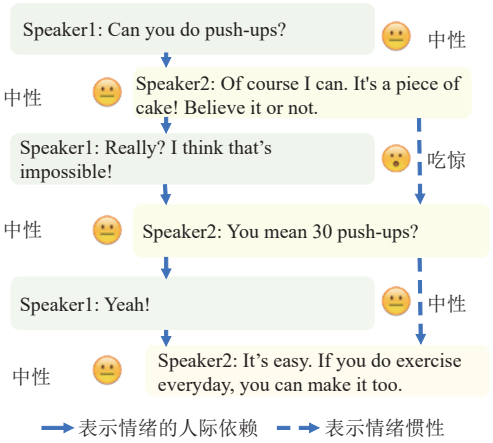


Fig. 1 Round 2 conversation of Dailydialog dataset

图 1 Dailydialog 数据集第 2 轮对话

第 3 个挑战是：如何提升模型识别相似情绪标签的能力. 对话情绪识别任务中情绪标签具有语义相似性，比如“兴奋”和“快乐”，“悲伤”和“沮丧”. Gao 等人[10] 提出的 SimCSE(simple contrastive learning of sentence embeddings)和 Yan 等人[11] 提出的 ConSERT (contrastive framework for self-supervised sentence representation transfer)证明了对比学习在自然语言处理任务中能捕捉到不同类别样本之间的区别. 本文引入对比学习任务帮助模型识别相似情绪标签，但传统对比学习构建的负样本与原有对话不具备上下文依赖关系，那么如何在不破坏对话上下文依赖关系的情况下捕捉不同类别样本之间的区别是应对挑战的关键.

针对现存挑战，本文基于 RoBERTa[12](robustly optimized BERT)设计了基于课程学习和知识图谱的对话情绪识别模型(curriculum learning and knowledge graphs for emotion recognition model in conversations, CK-ERC)，预训练阶段抽取含有常识和语法知识的结构化数据，在 K-Adapter[13](infusing knowledge into pre-trained models with adapters)工作的基础上预训练了常识编码层和语法编码层，为 CK-ERC 模型融入常识知识图谱和语法知识图谱；在数据处理阶段，实时抽取当前对话参与者的对话历史信息建立角色画像，建模对话上下文依赖关系；在微调阶段，引入监督对比学习任务来提升模型识别相似情绪标签的能力；在训练策略上，为了设计不破坏对话上下文依赖关系的课程学习策略，本文在 Superloss[14] 工作的基础上，设计了基于动态阈值的课程学习策略 DTS(dynamic threshold-superloss)，通过动态阈值调整训练样本权重，

优化模型性能.

本文的主要贡献有 3 点：

1)在基于预训练模型的对话情绪识别任务中，本文为模型融入常识和语法知识图谱，帮助模型准确建模口语化信息.

2)CK-ERC 模型在训练策略上设计了基于动态阈值的课程学习策略，按照文本丰富度从高到低、情绪转移频率从低到高的策略提升模型情绪识别能力；微调阶段引入监督对比学习任务来提升模型识别相似情绪标签的能力.

3)CK-ERC 模型在双人对话、多人对话、模拟对话和日常对话等多种对话模式下，显著优于其他模型，并在 MELD 和 EmoryNLP 数据集上获得最佳表现.

1 相关工作

1.1 对话领域情绪识别

对话领域情绪识别是基于对话的情感分类任务，需要建模对话上下文依赖，包括自我依赖和人际依赖. 自我依赖又称情绪惯性，表示对话参与者当前情绪和历史情绪之间的情绪依赖关系，人际依赖是不同对话参与者之间的情绪依赖关系. 通过建立对话上下文来建立情绪依赖关系，根据建模对话上下文信息的基线模型，将对话领域情绪识别模型分为 3 类.

1) 基于递归模型建模上下文信息. Hazarika 等人[15]提出的交互式对话记忆网络(interactive conversational memory network, ICON)、Hazarika 等人[16]提出的对话记忆网络(conversational memory network, CMN)、Jiao 等人[17]提出的分层门控的循环单元(hierarchical gated recurrent units, HIGRU)都通过门控递归单元(gated recurrent units, GRU)建模句子级别和对话级别的对话上下文信息，进而识别句子情绪标签. Majumder 等人[18]提出的对话循环神经网络模型 Dialoguernn(an attentive RNN for emotion detection in conversations)采用了全局 GRU 建模全局信息，并采用多个对话 GRU 建模对话参与者信息. 在多轮对话中该模型通过跟踪对话参与者的状态信息，建立对话中的人际依赖、分析对话参与者之间的情绪影响，帮助准确识别情绪.

2) 基于图神经网络建模上下文信息，并使用图神经网络来捕捉对话中的依赖关系. Ghosal 等人[19]提出的对话图卷积神经网络(dialogue graph convolutional network, DialogueGCN)使用 GRU 建模句子和对话级别的上下文信息，并将 GRU 编码的嵌入表示传入图神经网络，获得对话参与者之间的依赖关系. Ishiwatari 等人[20]提出的关系感知的图形注意网络(relation-aware graph attention networks, RGAT)使用图神经网络建模对话信息，并在对话图卷积神经网络基础上添加了关系位置编码. 在此基础上，Shen 等人[21]设计了有向无环图来建模对话中多种依赖关系. 彭韬等人[22]提出了基于依存关系的异构图卷积层来建模对话中多种依赖关系.

3) 基于预训练模型建模上下文信息. Li 等人[23]提出的 HiTrans(hierarchical transformer)用预训练模型 BERT[24](bidirectional transformer encoder)来建模句子级别上下文信息，通过 Transformer[25]模型来建模对话级别上下文信息、捕捉对话参与者之间的情绪交互. Shen 等人[26]提出的 DialogXL 对 XLNet[27]模型进行修改，将段级结构改为语篇级结构，可以建模更长的历史对话信息. 同时，在不同的注意力头上使用不同的掩码机制，使每个注意力头可以关注到不同的对话信息.

1.2 课程学习

课程学习是一种动态分配权重的训练策略，在初始阶段增大简单样本权重，随着训练的进行，逐渐提高困难样本的权重. 先对平滑的损失函数进行优化，逐步过渡到不平滑的损失函数，可以显著提高模型的性能. Bengio 等人[28]提出了课程学习策略，证明了课程学习在机器学习中的有效性，Hacohen 等人[29]进一步从数学角度分析了课程学习在深度学习中的有效性. 大量实验已经证明了好的课程学习策略能够提高模型精度并加快模型收敛. Zhou 等人[7]将课程学习应用到机器翻译任务上，提高了模型的精度和收敛速度；Huang 等人[8]将课程学习应用于关系提取任务上，提高了模型的精度和收敛速度.

1.3 预训练模型融入知识图谱

随着基于 BERT 的预训练模型在自然语言处理(NLP)相关的各种任务中取得了成功，为预训练模型融入特定领域的知识图谱，以获得特定领域的专业知识也成为了热点. Zhang 等人[30]提出的 ERNIE 模型使用 TransE[31](translating embeddings)知识嵌入算法对维基数据中的三元组进行训练获得实体嵌入，将知识图谱融入预训练模型中. 同时，优化 BERT 模型架构解决异构向量难以融合的问题. Lauscher 等人[32]提出的 LIBERT(lexically informed BERT)为词汇网络(WordNet)注入同义词和下义词关系的知识，使用特殊字符分隔词对，预测词对之间是否有指定关系. Xiong 等人[33]提出的弱监督知识预训练模型将输入数据中的实体替换成相同类型的实体，并通过让模

型判断实体是否为随机替换的，为预训练模型融入知识图谱.

1.4 对比学习

对比学习将一个样本的不同的、增强过的新样本在嵌入空间中的距离拉近，然后和其他不同的样本之间的距离拉大. 对比学习可以捕捉到同类样本之间的相同特征，并捕捉到不同类样本之间的区别，进而提升分类模型的分类准确率. Khosla 等人[34]提出的对比学习方法在分类任务中使用不同的正负样本构造策略：将训练数据中的标签应用到基于自监督训练的监督对比学习中，将批次中相同标签的样本作为正样本、不同标签的样本作为负样本. 这种策略高效地利用了标签信息，但使用交叉熵损失可能会使模型收敛到较差的局部最优点，导致模型性能下降. 为了解决上述问题，Gunel 等人[35]提出了监督对比学习，在模型微调阶段使用监督对比学习，利用三元组损失函数降低交叉熵损失带来的干扰.

2 CK-ERC 模型

2.1 问题定义

多轮对话数据集表示为C，对话参与者集合表示为S，情绪标签集合表示为Emo. n表示本轮对话中有n个句子，m表示当前句子中含有m个单词. 每轮对话都由一系列的句子$u_1, u_2, \cdots, u_n$构成，其中u_n表示本轮对话中的第n个句子；每个句子都有对应的情绪标签$y_1, y_2, \cdots, y_n$，其中y_n表示本轮对话中的第n个句子的情绪标签；每个句子都由若干个单词$w_1, w_2, \cdots, w_m$构成，其中w_m表示本句对话中的第m个单词；本文关注实时对话，模型输入为最近k轮的对话和对话参与者$s_{i-k}, u_{i-k}, \cdots, s_i, u_i$，其中当前时刻对话和对话参与者分别用$u_i$和$s_i$表示，任务目标是实时预测当前句子$u_i$的情绪标签.

2.2 模型概述

本节从数据构造、特征编码模块和多任务学习模块 3 方面对 CK-ERC 模型概述，模型架构如图 2 所示.

对话情绪识别受到情绪惯性和情绪依赖的影响. 对话情绪识别存在情绪惯性，即当前对话参与者的情绪受之前情绪的影响. 本文实时抽取当前对话参与者的历史对话信息，建立情绪惯性. 情绪具有人际依赖，即当前对话参与者的情绪会受到其他对话参与者的影响. 本文实时抽取当前对话的前k轮对话，并将对话参与者与对话拼接，建立人际依赖关系. 并设置提示信息，使模型适应下游任务，如式(1)所示：

$$data_i = \left[history, s_{i-k}, u_{i-k}, \cdots, s_i, u_i, question\right]. \tag{1}$$

$s_{i-k}, u_{i-k}, \cdots, s_i, u_i$表示最近$k$轮的对话参与者和对

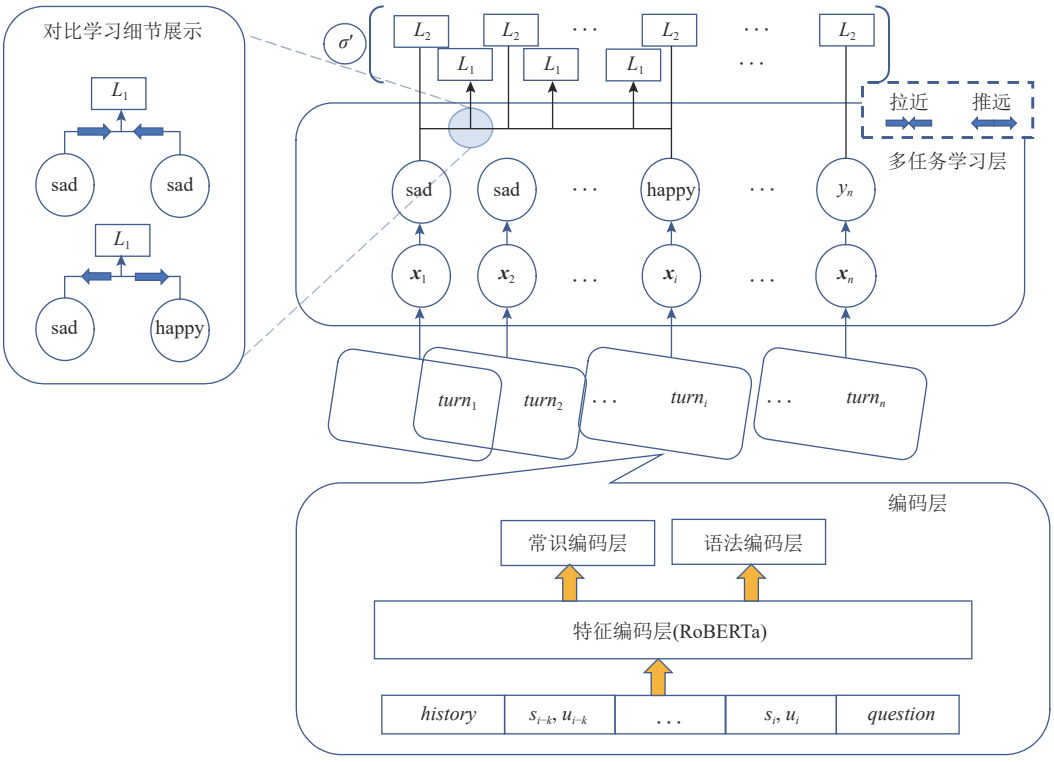


Fig. 2 Architecture diagram of a conversational emotion recognition model based on knowledge graph and curriculum learning

图 2 基于知识图谱和课程学习的对话情绪识别模型架构图

话；*history*表示实时抽取的对话历史信息；*question*表示提示信息“how does s_i feel now?”用来提示模型需要识别当前对话参与者的情绪.

特征编码模块通过 RoBERTa 预训练模型获得输入数据嵌入表示. 通过关系分类任务和依存关系分析任务并行训练 Factual Adapter 和 Linguistic Adapter 模型，为数据嵌入表示融入常识知识和语法知识，帮助模型识别对话中具有二义性的词语和省略句.

多任务学习模块由对比学习任务（任务 1）和建模情绪序列任务（任务 2）构成：任务 1，通过监督对比学习将一个批次中不同类型情绪标签之间的距离拉大，进而帮助模型区分相似的情绪标签，损失函数是 L_1；任务 2，通过条件随机场（conditional random field，CRF）对情绪序列建模，捕捉对话中的情绪转移概率，损失函数是 L_2. 在学习策略上，使用本文设计的 DTS 实现课程学习策略，设置动态阈值使模型按照由易到难的顺序进行学习优化，为损失函数分配的权重记为 σ'.

2.3 编码层

2.3.1 特征编码层

RoBERTa 模型是基于 BERT 模型改进得到的预训练模型，RoBERTa 是基于多头自注意力机制的模型，可以用于挖掘数据文本的语法和句法结构，并能够充分地学习到数据文本中字级、单词级和句子级的特征表示. RoBERTa 由多层 Transformer 的编码器构成，Transformer 的编码器主要由多头自注意力层（multi-head self-attention layer）、线性变换和残差连接子层构成. 根据表 1 可知，RoBERTa 由 12 层 Transformer 编码器构成，具体结构如图 3 所示.

Table 1 Parameter Settings of RoBERTa Model

表 1 RoBERTa 模型的参数设置

模型	层数	隐藏尺寸	每层头数
RoBERTa-large	24	1 024	16
RoBERTa	12	768	12

在模型结构上 RoBERTa 和 BERT 是一致的，RoBERTa 相比于 BERT 主要有 3 点改变：第 1 点是 RoBERTa 在预训练阶段去掉了下一句预测任务. 第 2 点是 RoBERTa 使用动态掩码策略，每次向 RoBERTa 模型输入序列时都会生成新的掩码模式. 随着训练数据不断输入变化掩码，RoBERTa 模型会逐渐适应不同的掩码策略，学习到不同的语言表征. 第 3 点是 RoBERTa 在预训练阶段，使用更大的批次和学习率进行训练，并使用更多的训练数据进行训练. 因此，

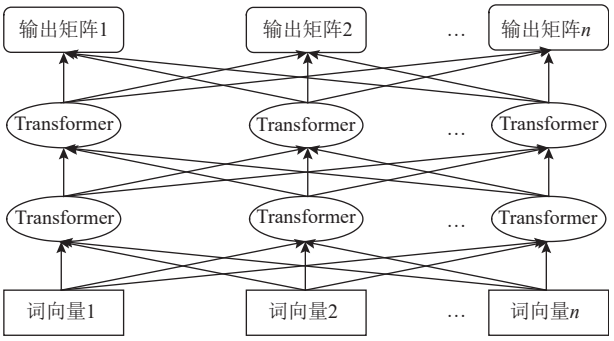


Fig. 3 RoBERTa model architecture

图 3 RoBERTa 模型架构

采用 RoBERTa 作为特征编码层可以帮助数据获得更丰富的嵌入表征.

模型 RoBERTa 的计算结构用 $Embedding(\cdot)$ 表示，将输入数据 $data_i$ 送入 RoBERTa 模型获得嵌入表示 $\boldsymbol{v}_i$ 的过程为

$$\boldsymbol{v}_i = Embedding(data_i). \tag{2}$$

2.3.2 知识编码层

知识编码层将常识知识图谱和语法知识图谱融入到对话情绪识别模型中，帮助模型识别对话中隐含的常识和语法信息. 如图 4 所示，本文通过 Factual Adapter 和 Linguistic Adapter 为情感编码融入常识知识和语法知识，将 Factual Adapter 和 Linguistic Adapter 连接到 RoBERTa 中不同的 Transformer 层上，为预训练模型融入常识知识图谱和语法知识图谱.

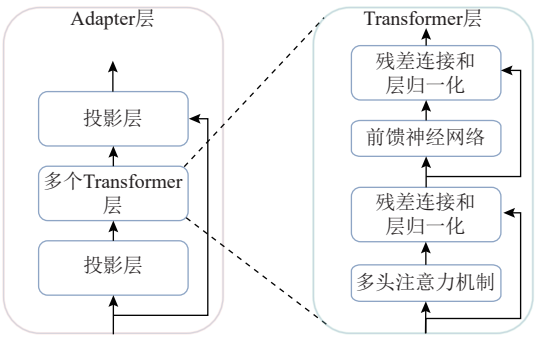


Fig. 4 Structure of Adapter layer

图 4 Adapter 层结构

K-Adapter 支持持续学习，在为模型架构引入新知识时不会引起灾难性遗忘. 传统模型引入新知识时，主干模型需要参与更新，用旧知识预训练好的参数会被新知识改变；K-Adapter 引入多种知识时，采用平行训练各个 Adapter 网络的策略. K-Adapter 包含多个 Adapter 层，每层 Adapter 由多个 Transformer 层和 2 个投影层构成，并且在 2 层投影层之间添加残差处理. 预训练好的 K-Adapter 连接在 RoBERTa 指定层后

面, 指定层记作a, b.

获得新知识的流程为: 第 1 步, 第a层 RoBERTa 对应的 Adapter 层记作$function_1$, 将第a层 RoBERTa 输出隐藏状态$\boldsymbol{r}_1$送入$function_1$, 输出的结果记作$\boldsymbol{out}_1$; 第 2 步, 第b层 RoBERTa 对应的 Adapter 层记作$function_2$, 将第b层 RoBERTa 输出隐藏状态$\boldsymbol{r}_2$和上一层输出的$\boldsymbol{out}_1$拼接后送入$function_2$, 输出记作$\boldsymbol{out}_2$; 第 3 步, 将$\boldsymbol{out}_2$与 RoBERTa 最后一层的输出隐藏状态$\boldsymbol{r}_3$拼接即可得到融入常识和语法知识的嵌入表示$\boldsymbol{out}_3$.具体流程为:

$$\boldsymbol{out}_1 = function_1(\boldsymbol{r}_1), \tag{3}$$

$$\boldsymbol{out}_2 = function_2(\boldsymbol{r}_2 + \boldsymbol{out}_1), \tag{4}$$

$$\boldsymbol{out}_3 = r_3 + \boldsymbol{out}_2. \tag{5}$$

使用结构化数据预训练 K-Adapter 模型, 为模型融入常识和语法知识, 预训练过程为: 使用常识知识数据集 T-Rex 为 Adapter 融入常识知识图谱, T-Rex 是维基百科摘要和维基数据三元组之间的大规模对齐数据集. 模型根据上下文对给定实体对的关系标签进行分类, 也就是对结构化常识数据进行关系分类; 通过 Standford Parser 处理 BookCorpus 数据集, 构建结构化语法数据集, 进行依存关系分类任务, 预测给定句子中每个词语在依存分析结果中的父节点, 进而获得语法知识. 将融入常识知识图谱的 Adapter 得到的嵌入表示记作$\boldsymbol{Fout}$, 将融入语法知识图谱的 Adapter 得到的嵌入表示记作$\boldsymbol{Lout}$, 融入语法知识和常识知识的嵌入表示记作$\boldsymbol{x}_i$:

$$\boldsymbol{x}_i = \boldsymbol{Lout} + \boldsymbol{Fout}. \tag{6}$$

经过全连接层得到本回合所有对话的情绪预测概率值$\boldsymbol{p}_i$.

2.4 多任务学习层

2.4.1 监督对比学习

在监督训练阶段, 本文采用有监督对比学习, 将批次中具有相同标签的所有样例都视为正样本, 将同一批次中标签不同的样例作为负样本. 在嵌入空间中, 监督对比学习任务将具有不同情绪标签样本之间的余弦距离拉大, 效果如图 5 所示.

监督对比学习将样本在嵌入空间的余弦距离加入到优化目标函数中. 随着模型优化, 嵌入空间内部不同类别的标签之间的余弦距离被拉大, 消除语义相似性带来的影响, 进而帮助模型区分相似情绪, 提升模型情绪识别的表现. 式(7)为有监督微调阶段的优化目标:

$$L_1 = (1-\delta)L_{\mathrm{CE}} + \delta L_{\mathrm{SCL}}. \tag{7}$$

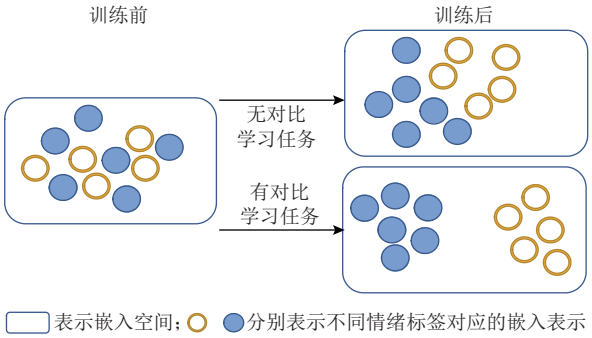


Fig. 5 Demonstration of the effects of supervised comparative learning

图 5 监督对比学习效果展示

CK-ERC 模型的优化目标由 2 部分构成: L_{SCL}表示有监督对比学习 (supervised contrastive learning, SCL) 损失, L_{CE}表示交叉熵 (cross entropy, CE) 损失, δ是一个标量加权超参数, 分配二者的权重. L_{SCL}有监督对比学习损失函数如式(8)~(10)所示:

$$L_{\mathrm{SCL}} = -\frac{1}{M-1}\sum_{i=1}^{N}\sum_{j=1}^{N}1_{i\neq j}1_{y_i=y_j}\ln\left(\frac{J}{K}\right), \tag{8}$$

$$K = \sum_{k=1}^{N}1_{i\neq k}\exp\left(\left(\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j\right)/\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|\right)/\xi\right), \tag{9}$$

$$J = \exp\left(\left(\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j\right)/\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|\right)/\xi\right). \tag{10}$$

在式(8)~(10)中, N表示每批次中训练样本的数量, M表示在当前批次中与y_i有相同标签的样例数量, 其中1 表示指示函数, 只有满足指定条件时才取值为 1, 例如$1_{i\neq k}$只有满足条件$i\neq k$时才取值为 1. $\boldsymbol{x}_i$表示数据$data_i$经过情感编码后得到的嵌入表示, y_i表示嵌入表示$\boldsymbol{x}_i$对应的标签, ξ是可调节的标量温度参数, 用于控制类的分离程度. L_{CE}交叉熵损失函数为

$$L_{\mathrm{CE}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{e=1}^{E}y_{ie}\times\ln(p_{ie}), \tag{11}$$

其中, N表示每批次中训练样本的数量, E表示数据集中情绪类别数量. y_{ie}取值 0 或 1, 第i个样本的真实类别等于e时取值为 1, 否则取值为 0. p_{ie}表示第i个样本属于类别e的预测概率.

2.4.2 情绪序列建模

对话参与者的情绪会受到彼此的影响, 并受到情绪惯性的影响. 情绪在对话参与者之间并不是随机传递转移的, 对话的情绪具有明显的时间相关性, 转移具有规律性. 为此, 使用条件随机场来捕捉情绪的顺序信息. 条件随机场是一种无向图模型, 同时结合了最大熵模型和隐马尔可夫模型的特点, 常用于对顺序依赖性建模. 从语料编码器中获得对话参与

者的情绪概率分布后，我们将其输入到 CRF 层，学习对话层面的情绪传播规律：

$$\max(\boldsymbol{p}_i|ALL), \tag{12}$$

其中，$ALL=(\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n)$，表示所有可能的情绪序列，在所有可能的情绪序列中最大化真实情绪序列的概率，达到优化模型的效果. 优化目标函数为

$$L_{\mathrm{CRF}}(\boldsymbol{\theta}) = -\ln(\boldsymbol{p}), \tag{13}$$

其中$\boldsymbol{\theta}$是表示条件随机场中可训练参数，$\boldsymbol{p}$表示出现当前情绪序列的概率.

2.5 DTS 课程学习框架

课程学习是一种训练策略，通过模拟人类学习过程，使模型先从简单的样本开始学习，逐步进阶到困难样本，可以显著提升模型性能. 课程学习策略已经应用在计算机视觉和自然语言处理等多种应用场景中，提高模型的泛化能力收敛率. 课程学习的核心任务是按照学习优先（learning priority）程度重新排列训练样本输入模型的顺序. 学习优先程度需要由难度测量器（difficulty measurer）判断，进而通过训练调度器（training scheduler）重排训练样本输入模型的顺序.

对话情绪识别任务的数据集具有严格的上下文依赖关系，无法使用传统的课程学习策略. 为此，本文设计了 DTS 课程学习框架提升模型性能，通过动态阈值来动态分配权重实现课程学习策略，防止模型陷入较差的局部最优点. DTS 课程学习策略在初始阶段阈值较低，为简单样本分配较高的权重. 随着训练的进行，阈值逐渐提高，提高了困难样本的权重. 维持低于阈值的简单样本权重不变，降低高于阈值样本的权重，关于 DTS 课程学习策略具体公式为：

$$L_\lambda(L_i, \sigma_i) = (L_i - \tau')\sigma_i + \lambda(\ln\sigma_i)^2, \tag{14}$$

$$\tau' = \tau + epoch_i \times \vartheta. \tag{15}$$

式（14）为 DTS 的表达式，用来降低高于阈值样本的权重，任务损失记作L_i和动态阈值记作τ'，低于τ'的样本被判断为简单样例，高于τ'的样本被判断为困难样例，需要降低权重. τ为阈值初始值，ϑ为阈值增量，$epoch_i$表示训练进行到第i轮动态阈值随训练升高，通过$\tau + epoch_i \times \vartheta$来实现动态阈值变化. σ_i为置信度，用来调节输入样本对模型更新的贡献大小. $\lambda(\ln\sigma_i)^2$为正则化项，λ为正则化超参数.

实验证明将σ_i设置为超参数会导致模型收敛变慢，因此选取σ_i极限处的收敛值作为权重，$\sigma_i'(L_i)$表示σ_i求极限的公式，具体公式为

$$\sigma_i'(L_i) = \arg\min(L_\lambda(L_i, \sigma_i)). \tag{16}$$

通过数学公式推导得到σ_i的表达式，其中$LamW$表示朗伯 W 函数（Lambert W function），β用来避免无限值的出现，σ_i的表达式为

$$\sigma_i = \exp\left(-LamW\left(\frac{1}{2}\max\left(-\frac{2}{\mathrm{e}}, \beta\right)\right)\right), \tag{17}$$

$$\beta = \frac{l_i - \tau'}{\lambda}. \tag{18}$$

2.6 训练与预测

在训练阶段，实时抽取当前对话参与者的对话历史信息以及前 3 轮对话，构建数据$data_i$，通过特征编码层和情感编码层获得融入语法知识和常识知识的嵌入表示$\boldsymbol{x}_i$. 将$\boldsymbol{x}_i$送入到多任务学习层，在监督对比学习任务中优化目标函数为L_1，将同一批次中具有相同标签的嵌入表示在嵌入空间内的余弦距离作为优化目标；在情绪序列建模任务中优化目标函数为L_2，在所有可能的情绪序列中最大化真实情绪序列的概率，总体损失函数表示为

$$L = L_1 + L_2. \tag{19}$$

DailyDialog 数据集中超过 82% 的情绪标签是中性（neutral），导致情绪标签分布不均匀，基于 DailyDialog 数据集实验的对话情绪识别模型，都不预测中性标签. 为了对比试验的公正性，本文也不预测中性标签. 由于条件随机场在训练阶段捕捉到的情绪变化规律与中性标签密切相关，进而会在预测阶段干扰模型，所以本文基于 DailyDialog 数据集实验使用归一化指数函数 softmax 函数替代条件随机场.

在 MELD, EmoryNLP, IEMOCAP 数据集上，预测阶段使用条件随机场的 Viterbi 算法来解码最佳情绪序列. 在 DailyDialog 数据集上，取所有情绪概率中的最大值作为预测结果.

3 实　　验

3.1 数据集和评测指标

为了充分验证模型的有效性，本文在 4 个数据集上设计了 36 组对比实验和 32 组消融实验，这 4 个数据集分别是 MELD, EmoryNLP, IEMOCAP 和 DailyDialog，其中包括双人对话、多人对话、模拟对话和日常对话等多种对话模式.

IEMOCAP 是由 10 个对话参与者构成的数据集，情绪标签分别是“快乐”“悲伤”“愤怒”“兴奋”“沮丧”和“中立”. 训练数据集和验证数据集按照 9∶1 的比例进行分割. DailyDialog 数据集是由 2 个对话参与者模拟日常对话产生的数据集，情绪标签分别是“愤怒”“厌恶”“恐惧”“快乐”“惊讶”“悲伤”和“中性”. MELD 和 EmoryNLP 数据集都是从《Friend》这部电视

剧角色对话中提取的, MELD 的情绪标签列表是“愤怒”“厌恶”“悲伤”“喜悦”“惊讶”“恐惧”和“中立”. EmoryNLP 的情绪标签列表是“快乐”“和平”“强大”“恐惧”“疯狂”“悲伤”和“中性”, 数据集的相关信息如表 2 所示.本文实验评测指标均为 Weighted $F1$, Weighted $F1$ 通过对每类样本的 $F1$ 进行加权平均来计算整体性能, 适合数据样本类别不平衡的情况.

Table 2 Analysis of Datasets

表 2 数据集分析

数据集	情绪标签
MELD	愤怒、厌恶、悲伤、喜悦、惊讶、恐惧和中立
EmoryNLP	快乐、和平、强大、恐惧、疯狂、悲伤和中性
DailyDialog	愤怒、厌恶、恐惧、快乐、惊讶、悲伤和中性
IEMOCAP	快乐、悲伤、愤怒、兴奋、沮丧和中立

3.2 实验设置

本文实验基于 Pytorch 环境, 使用开源社区 HuggingFace 中的 RoBERTa 预训练模型作为基础模型, 优化器是 AdamW, 批次大小为 8, 预训练模型(LM 层)、全连接层(FC 层)和条件随机场(CRF 层)的学习率分别为 1E−5, 1E−4, 1E−3, RoBERTa-large 中 MELD 数据集的实验在 32 GB 显存 NVIDIA A100 GPU 上进行, 其余实验全部在 24 GB 显存的 NVIDIA 3090 GPU 上进行. CK-ERC(大)和 CK-ERC 分别表示基于 RoBERTa-large 和 RoBERTa-base 的 CK-ERC 模型.

模型表现最佳的超参数设置如表 3 所示, 在知识图谱列, 将常识编码器和语法编码器连接在多个 RoBERTa 层后面, 用 *list* 列表记录多个 RoBERTa 层; 对比学习的超参数 ξ,δ 分别表示可调节的标量温度参数和标量加权超参数; 课程学习超参数 τ,ϑ 分别表示阈值初始值和阈值增量; 句子长度 *length* 表示句子中的单词数量.

Table 3 Hyperparameter Setting

表 3 超参数设置

数据集	*list*	(ξ,δ)	(τ,ϑ)	*length*
MELD	[0,11,23]	(0.9,0.3)		256
EmoryNLP	[23]	(0.8,0.3)	(50,0.02)	230
DailyDialog	[23]	(0.58,0.3)	(38,0.003)	256
IEMOCAP	[23]	(0.8,0.3)	(9,0.000 03)	256

3.3 对比实验

本文提出的 CK-ERC 模型, 解决了对话情绪识别数据具有口语化、主题跨度大和标签具有语义相似性所带来的挑战. 为了进一步验证模型的优异性, 本文设计了 36 组对比实验. 将 CK-ERC 模型和 7 种对话情绪识别模型比较, 对比实验选用模型为覆盖本领域性能最优的模型和经典模型.

对比实验的基线模型有: Zhong 等人[36] 提出的知识增强 transformer (knowledge-enriched transformer, KET) 为对话情绪识别模型引入常识知识库和情感词典, Song 等人[37] 提出的 EmotionFlow 模型将 RoBERTa 用到对话情绪识别任务中, Sun 等人[38] 提出的 ERMC-DisGCN(a discourse-aware graph convolutional neural network)模型设计了一个关系卷积, 以利用对话者的自说自话的依赖性来传播语境信息, 以及 1.1 节介绍过的 RGAT 模型、DialogueGCN 模型、对话情绪识别常识引导框架(commonsense-guided framework for emotion identification in conversations, COSMIC)模型以及 HiTrans 模型.

对比实验结果如表 4 所示, CK-ERC(大) 和 CK-ERC 分别在 MELD, EmoryNLP, DailyDialog 这 3 个数据集取得最优结果和次优结果, CK-ERC(大) 在 IEMOCAP 数据集取得次优结果, 实验结果证明了 CK-ERC 模型性能的优异性. CK-ERC 模型在 MELD, EmoryNLP, DailyDialog, IEMOCAP 数据集上同时表现优异, 这 4 个数据集覆盖了日常对话、多人对话、双人对话和模拟对话等多种对话模式, 表明 CK-ERC 模型适用于多种对话场景, 更加贴近现实应用场景.

Table 4 Comparison of *F*1 on Four Datasets

表 4 4 个数据集上的 *F*1 对比 %

模型	MELD	EmoryNLP	DailyDialog	IEMOCAP
KET	58.18	34.39	53.37	59.56
DialogueGCN	58.39			**66.05***
RGAT	60.91	34.42	54.31	65.22
EmotionFlow	66.50			
ERMC-DisGCN	64.22	36.38		
COSMIC	65.21	38.11	58.48	65.28
HiTrans	61.94	36.75		64.50
CK-ERC (本文)	**65.64**	**38.58**	**60.29**	62.37
CK-ERC (大) (本文)	**67.20***	**41.17***	**61.80***	**65.82**

注: 将最优结果和次优结果设为黑体字, 并在最优结果后面标记*.

3.4 消融实验

为了证明 CK-ERC 模型的优越性, 本文设计了 32 组消融实验并对消融实验结果进行分析. 实验结果如表 5 所示. 为了进一步证明每种方法取得的提高是否针对相应的问题, 本文标记出口语化对话数据和具有语义相似标签的数据, 针对具体问题设计消融实验.

Table 5 Ablation Experiments on Four Datasets

表 5 4 个数据集的消融实验 %

模型	消融	MELD	EmoryNLP	DailyDialog	IEMOCAP
CK-ERC（本文）		65.64	38.58	60.29	62.37
	KG	65.62	37.87	60.25	62.26
	CL	65.29	37.26	60.07	61.88
	SCL	65.15	37.09	59.89	61.50
CK-ERC（大）（本文）		67.15	41.14	61.80	65.82
	CL	67.20	40.88	60.70	63.71
	KG	66.51	38.46	59.63	63.65
	SCL	66.07	38.09	59.45	62.81

注：SCL 表示对比学习任务，CL 表示课程学习策略，KG 表示知识编码层.

3.4.1 知识编码层有效性分析

消融知识编码层后，模型 Weighted *F*1 值有显著下降，在 CK-ERC 上模型 Weighted *F*1 值分别下降 0.02 个百分点、0.71 个百分点、0.04 个百分点、0.11 个百分点，在 CK-ERC(大)上模型 Weighted *F*1 值分别下降 0.69 个百分点、2.42 个百分点、1.07 个百分点、0.06 个百分点，进而证明了为对话情绪识别模型融入常识知识图谱和语法知识图谱的有效性. 为了进一步验证知识编码层是否能够帮助模型准确建模口语化信息，本文设计了如下实验：

在 MELD 测试集上，标记出 50 条含有二义性词语的语句和 50 条是省略句的语句. 利用 CK-ERC 模型和消融知识编码层后的 CK-ERC 模型分别对 MELD 测试集进行预测，记录上述 100 条数据索引对应的预测结果和答案标签. 实验结果显示，在含有二义性词语的语句上，消融知识编码层后的 CK-ERC 模型情绪识别准确率为 54%，CK-ERC 模型面向对话的情感识别和情感原因发现研究情绪识别准确率为 62%；在省略句上，消融知识编码层后的 CK-ERC 模型情绪识别准确率为 56%，CK-ERC 模型情绪识别准确率为 60%. 实验显示，在口语化对话数据上，知识编码层能够显著提高模型情绪识别准确率，进而证明融入常识和语法知识图谱的方法能够帮助模型准确建模口语化信息.

3.4.2 监督对比学习任务有效性分析

消融监督对比学习任务后，模型 Weighted *F*1 值显著下降，在 CK-ERC 上模型 Weighted *F*1 值分别下降 0.14 个百分点、0.17 个百分点、0.18 个百分点、0.33 个百分点，在 CK-ERC(大)上模型 Weighted *F*1 值分别下降 0.44 个百分点、0.37 个百分点、0.18 个百分点、0.84 个百分点，进而证明了监督对比学习任务的有效性. 为了进一步验证监督对比学习任务是否能帮助模型区分相似情绪标签，本文设计了如下实验：

在 EmoryNLP 测试集中，有 288 条测试数据样本的情绪标签是中性的(neutral)，有 111 条测试数据样本的情绪标签是平静的(peaceful)，neutral 和 peaceful 具有语义相似性. 统计把 neutral 预测成 peaceful 的数量和把 peaceful 预测成 neutral 的数量，neutral 和 peaceful 相互预测错误的数量可以用来表示 neutral 和 peaceful 的语义区分度. 利用 CK-ERC 模型和消融监督对比学习任务后的 CK-ERC 模型分别对 EmoryNLP 测试集进行预测. 消融监督对比学习任务后的 CK-ERC 模型把 neutral 预测成 peaceful 的数量是 47 个、把 peaceful 预测成 neutral 的数量是 29 个. CK-ERC 模型把 neutral 预测成 peaceful 的数量是 28 个、把 peaceful 预测成 neutral 的数量是 16 个. 实验结果显示，监督对比学习任务能够帮模型区分具有语义相似性的情绪标签. 监督对比学习将样本在嵌入空间的余弦距离加入到损失函数中，随着模型优化，嵌入空间中不同类别的样本之间的余弦距离被拉大，进而消除情绪标签相似性带来的影响.

3.4.3 课程学习策略有效性分析

消融课程学习策略后，模型性能有显著下降，在 CK-ERC 上模型 Weighted *F*1 值分别下降 0.33 个百分点、0.61 个百分点、0.18 个百分点、0.38 个百分点，在 CK-ERC(大)上模型 Weighted *F*1 值分别下降−0.05 个百分点、0.29 个百分点、0.90 个百分点、2.11 个百分点，进而证明了基于动态阈值的课程学习策略的有效性. 模型 Weighted *F*1 值下降−0.05 个百分点(即上升 0.05 个百分点)的原因是，CK-ERC(大)基于 RoBERTa-large 模型融入了常识知识图谱和语法知识图谱，显存占用较大、模型训练时间较长，所以并未对 CK-ERC(大)模型中的课程学习超参数调节至最优.

将文本丰富度高、情绪转移频率低的训练样本称为简单样本，反之称为困难样本. 课程学习策略在训练初期赋予困难样本低权重，专注于学习简单样本，随着模型优化，逐步提高赋予困难样本的权重，这种训练策略针对性地解决了主题跨度大导致模型性能下降的问题.

为了进一步证明动态阈值的必要性，对动态阈值进行消融，设置阈值增量为 0，其他参数保持不变. CK-ERC 在 MELD, EmoryNLP, DailyDialog, IEMOCAP 数据集上的 Weighted *F*1 值分别为 65.03%, 37.89%, 59.76%, 60.67%，模型 Weighted *F*1 值分别下降了 0.61 个百分点、0.69 个百分点、0.53 个百分点、1.70 个百分点，这充分证明了动态阈值的有效性.

4 结 论

本文深入研究了对话数据集的特点，发现对话领域情绪识别任务中存在如下问题：对话文本具有口语化的特点导致模型难以准确建模语义信息，对话主题跨度大导致模型性能下降，模型难以识别具有语义相似性的情绪标签. 为解决上述问题，本文提出 CK-ERC 模型，在预训练阶段为模型融入常识和语法知识帮助模型建模语义信息，在微调阶段引入监督对比学习任务帮助模型识别相似情绪标签，在训练策略上设计了全新的课程学习策略优化模型性能. CK-ERC 模型在双人对话、多人对话、模拟对话和日常对话等多种对话模式下显著优于其他模型，在 MELD 和 EmoryNLP 数据集上获得最佳表现.

作者贡献声明：杜金明负责模型构思、代码运行编写、实验设计、实验结果分析、论文撰写与校对；孙媛媛负责模型构思、实验结果分析，并对论文撰写进行全程指导；林鸿飞、杨亮对论文撰写进行全程指导.

参 考 文 献

[1] Zhao Yanyan, Qin Bing, Liu Ting. Sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834−1848(in Chinese)
(赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834−1848)

[2] Li Ran, Lin Zheng, Lin Hailun, et al. Text emotion analysis: A survey[J]. Journal of Computer Research and Development, 2018, 55(1): 30−52(in Chinese)
(李然, 林政, 林海伦, 等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30−52)

[3] Yin Qingyu, Zhang Weinan, Zhang Yu, et al. A joint model for ellipsis identification and recovery [J]. Journal of Computer Research and Development, 2015, 52(11): 2460−2467(in Chinese)
(尹庆宇, 张伟男, 张宇, 等. 省略识别及恢复联合模型研究[J]. 计算机研究与发展, 2015, 52(11): 2460−2467)

[4] Ghosal D, Majumder N, Gelbukh A, et al. Cosmic: Commonsense knowledge for emotion identification in conversations [J]. arXiv preprint, arXiv: 2010.02795, 2020

[5] Zhang Duzhen, Chen Xiuyi, Xu Shuang, et al. Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer [C/OL] //Proc of the 28th Int Conf on Computational Linguistics. [2023-06-16].https://aclanthology.org/2020.coling-main.392

[6] Zhang Xiaoyu, Li Dongdong, Ren Pengjie, et al. Memory networks based knowledge-aware medical dialogue generation[J]. Journal of Computer Research and Development, 2022, 59(12): 2889−2900(in Chinese)
(张晓宇, 李冬冬, 任鹏杰, 等. 基于记忆网络的知识感知医疗对话生成[J]. 计算机研究与发展, 2022, 59(12): 2889−2900)

[7] Zhou Yikai, Yang Baosong, Wong D F, et al. Uncertainty-aware curriculum learning for neural machine translation [C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 6934−6944

[8] Huang Yuyun, Du Jinhua. Self-attention enhanced CNNs and collaborative curriculum learning for distantly supervised relation extraction [C] //Proc of the 24th Conf on Empirical Methods in Natural Language Processing and the 9th Int Joint Conf on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg, PA: ACL, 2019: 389−398

[9] Zeng Weixin, Zhao Xiang, Tang Jiuyang, et al. Iterative entity alignment via re-ranking[J]. Journal of Computer Research and Development, 2020, 57(7): 1460−1471(in Chinese)
(曾维新, 赵翔, 唐九阳, 等. 基于重排序的迭代式实体对齐[J]. 计算机研究与发展, 2020, 57(7): 1460−1471)

[10] Gao Tianyu, Yao Xingcheng, Chen Danqi. SimCSE: Simple contrastive learning of sentence embeddings [J]. arXiv preprint, arXiv: 2104.08821, 2021

[11] Yan Yuanmeng, Li Rumei, Wang Sirui, et al. ConSERT: A contrastive framework for self-supervised sentence representation transfer [J]. arXiv preprint, arXiv: 2105.11741, 2021

[12] Liu Yinhan, Ott M, Goyal N, et al. RoBERTa: A robustly optimized bert pretraining approach [J]. arXiv preprint, arXiv: 1907.11692, 2019

[13] Wang Ruize, Tang Duyu, Duan Nan, et al. K-Adapter: Infusing knowledge into pre-trained models with adapters [J]. arXiv preprint, arXiv: 2002. 01808, 2020

[14] Castells T, Weinzaepfel P, Revaud J. Superloss: A generic loss for robust curriculum learning [C/OL] // Proc of the 33rd Conf on Neural Information Processing Systems. [2023-06-16].https://proceedings.neurips.cc/paper/2020/hash/2cfa8f9e50e0f510ede9d12338a5f564-Abstract.html

[15] Hazarika D, Poria S, Mihalcea R, et al. ICON: Interactive conversational memory network for multimodal emotion detection [C] //Proc of the 23rd Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 2594−2604

[16] Hazarika D, Poria S, Zadeh A, et al. Conversational memory network for emotion recognition in dyadic dialogue videos [C] //Proc of the 56th Conf Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 2018−2122

[17] Jiao Wenxiang, Yang Haiqin, King I, et al. Higru: Hierarchical gated recurrent units for utterance-level emotion recognition [J]. arXiv preprint, arXiv: 1904.04446, 2019

[18] Majumder N, Poria S, Hazarika D, et al. Dialoguernn: An attentive RNN for emotion detection in conversations [C] //Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 6818−6825

[19] Ghosal D, Majumder N, Poria S, et al. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation [J]. arXiv preprint, arXiv: 1908.11540, 2019

[20] Ishiwatari T, Yasuda Y, Miyazaki T, et al. Relation-aware graph attention networks with relational position encodings for emotion

recognition in conversations［C］//Proc of the 25th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 7360−7370

［21］ Shen Weizhou, Wu Siyue, Yang Yunyi, et al. Directed acyclic graph network for conversational emotion recognition［J］. arXiv preprint, arXiv: 2105.12907, 2021

［22］ Peng Tao, Yang Liang, Sang Zhongyi, et al. Conversational sentiment analysis based on heterogeneous bipartite graphs［J］. Journal of Chinese Information Processing, 2021, 35(11): 135−142(in Chinese)
(彭韬, 杨亮, 桑钟屹, 等. 基于异构二部图的对话情感分析[J]. 中文信息学报, 2021, 35(11): 135−142)

［23］ Li Jingye, Ji Donghong, Li Fei, et al. HiTrans: A transformer-based context-and speaker-sensitive model for emotion detection in conversations［C/OL］//Proc of the 28th Int Conf on Computational Linguistics. ［2023-06-16］.https://aclanthology.org/2020.coling-main.370

［24］ Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding［J］. arXiv preprint, arXiv: 1810.04805, 2018

［25］ Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need［J］. arXiv preprint, arXiv: 1706.03762, 2017

［26］ Shen Weizhou, Chen Junqing, Quan Xiaojun, et al. DialogXL: All-in-one XLnet for multi-party conversation emotion recognition［C］//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 13789−13797

［27］ Yang Zhilin, Dai Zihang, Yang Yiming, et al. XLnet: Generalized autoregressive pretraining for language understanding［J］. arXiv preprint, arXiv: 1906.08237, 2019

［28］ Bengio Y, Louradour J, Collobert R, et al. Curriculum learning［C/OL］//Proc of the 26th Annual Int Conf on Machine Learning. ［2023-06-16］.https://dl.acm.org/doi/10.1145/1553374.1553380

［29］ Hacohen G, Weinshall D. On the power of curriculum learning in training deep networks［J］. arXiv preprint, arXiv: 1904.03626, 2019

［30］ Zhang Zhengyan, Han Xu, Liu Zhiyuan, et al. ERNIE: Enhanced language representation with informative entities［J］. arXiv preprint, arXiv: 1905.07129, 2019

［31］ Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data［C/OL］// Proc of the 26th Conf on Neural Information Processing Systems. ［2023-06-16］.https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html

［32］ Lauscher A, Vulić I, Ponti E M, et al. Informing unsupervised pretraining with external linguistic knowledge［J］. arXiv preprint, arXiv: 1909.02339, 2019

［33］ Xiong Wenhan, Du Jingfei, Wang W Y, et al. Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model［J］. arXiv preprint, arXiv: 1912.09637, 2019

［34］ Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning［C/OL］// Proc of the 33rd Conf on Neural Information Processing Systems. ［2023-06-16］. https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html

［35］ Gunel B, Du Jingfei, Conneau A, et al. Supervised contrastive learning for pre-trained language model fine-tuning［J］. arXiv preprint, arXiv: 2011.01403, 2020

［36］ Zhong Peixiang, Wang Di, Miao Chunyan. Knowledge-enriched transformer for emotion detection in textual conversations［J］. arXiv preprint, arXiv: 1909.10681, 2019

［37］ Song Xiaohui, Zang Liangjun, Zhang Rong, et al. EmotionFlow: Capture the dialogue level emotion transitions［C］// Proc of the 47th Int Conf on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2022: 8542−8546

［38］ Sun Yang, Yu Nan, Fu Guohong. A discourse-aware graph neural network for emotion recognition in multi-party conversation［C］// Proc of the 26th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 2949−2958



Du Jinming, born in 2000. Master. His main research interests include deep learning, natural language processing, and sentiment analysis.
杜金明, 2000 年生. 硕士. 主要研究方向为深度学习、自然语言处理、情感分析.



Sun Yuanyuan, born in 1979. PhD, professor, PhD supervisor. Senior member of CCF. Her main research interests include deep learning and natural language processing.
孙媛媛, 1979 年生. 博士, 教授, 博士生导师. CCF 高级会员. 主要研究方向为深度学习、自然语言处理.



Lin Hongfei, born in 1962. PhD, professor, PhD supervisor. Member of CCF. His main research interests include deep learning, natural language processing, and sentiment analysis. (hflin@dlut.edu.cn)
林鸿飞, 1962 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为深度学习、自然语言处理、情感分析.



Yang Liang, born in 1986. PhD, lecturer. His main research interests include deep learning, natural language processing, and information retrieval. (liang@dlut.edu.cn)
杨　亮, 1986 年生. 博士, 讲师. 主要研究方向为深度学习、自然语言处理、信息检索.