

属性建模与课程学习相结合的属性级情感分类方法*

叶 静^{1,2}, 向 露^{1,2}, 宗成庆^{1,2}

¹(中国科学院大学 人工智能学院, 北京 100049)

²(模式识别国家重点实验室 (中国科学院 自动化研究所), 北京 100190)

通信作者: 宗成庆, E-mail: cqzong@nlpr.ia.ac.cn



摘 要: 属性级情感分类任务旨在判断句子针对给定属性的情感极性, 因其广泛应用而备受关注. 该任务的关键在于识别给定属性相关的上下文描述, 并根据上下文内容判断发文者针对相应属性的情感倾向. 统计发现, 大约 30% 的评论中并不包含关于给定属性的明确情感描述, 但仍然传达了清晰的情感倾向, 这被称为隐式情感表达. 近年来, 基于注意力机制的神经网络方法在情感分析中得到了成功应用. 但该方法只能捕捉属性相关的显式情感描述, 而缺乏对隐含情感的有效分析和挖掘, 且往往将属性词与句子上下文分别建模, 使得属性词的表示缺乏上下文语义. 针对以上两个问题, 提出一种交叉融合属性局部和句子全局上下文信息的属性级情感分类方法, 并根据隐式和显式情感表达句子不同的分类难度采用课程学习提高模型的性能. 实验表明, 所提方法不仅对显式情感表达句子的属性情感倾向识别准确率高, 而且能够有效学习隐式情感表达句子的情感类别.

关键词: 属性级情感分析; 隐式情感分析; 注意力机制; 课程学习

中图法分类号: TP18

中文引用格式: 叶静, 向露, 宗成庆. 属性建模与课程学习相结合的属性级情感分类方法. 软件学报. <http://www.jos.org.cn/1000-9825/6963.htm>

英文引用格式: Ye J, Xiang L, Zong CQ. Aspect-level Sentiment Classification Combining Aspect Modeling and Curriculum Learning. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/6963.htm>

Aspect-level Sentiment Classification Combining Aspect Modeling and Curriculum Learning

YE Jing^{1,2}, XIANG Lu^{1,2}, ZONG Cheng-Qing^{1,2}

¹(School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China)

²(National Laboratory of Pattern Recognition (Institute of Automation, Chinese Academy of Sciences), Beijing 100190, China)

Abstract: Aspect-level sentiment classification task, which aims to determine the sentiment polarity of a given aspect, has attracted increasing attention due to its broad applications. The key to this task is to identify contextual descriptions relevant to the given aspect and predict the aspect-related sentiment orientation of the author according to the context. Statistically, it is found that close to 30% of reviews convey a clear sentiment orientation without any explicit sentiment description of the given aspect, which is called implicit sentiment expression. Recent attention mechanism-based neural network methods have gained great achievement in sentiment analysis. However, this kind of method can only capture explicit aspect-related sentiment descriptions but fails to effectively explore and analyze implicit sentiment, and it often models aspect words and sentence contexts separately, which makes the expression of aspect words lack contextual semantics. To solve the above two problems, this study proposes an aspect-level sentiment classification method that integrates local aspect information and global sentence context information and improves the classification performance of the model by curriculum learning according to different classification difficulties of implicit and explicit sentiment sentences. Experimental results show that the proposed method not only has a high accuracy in identifying the aspect-related sentiment orientation of explicit sentiment sentences but also can effectively learn the sentiment categories of implicit sentiment sentences.

Key words: aspect-level sentiment analysis; implicit sentiment analysis; attention mechanism; curriculum learning

* 收稿时间: 2022-09-14; 修改时间: 2022-11-03, 2023-03-24; 采用时间: 2023-05-05

随着网络用户生成内容的激增,从用户主观性评论文本中挖掘情感和观点的研究迅速发展. 早期的情感分析方法主要聚焦于判断文本的整体情感倾向,而在实际应用中细粒度的情感分析更为重要. 属性级情感分类 (aspect level sentiment analysis, ALSA) 是一种细粒度的情感分类任务,其旨在判断评论针对给定属性的情感倾向^[1,2]. 如评论“Great food, but the service was dreadful!”中,发作者针对“food”和“service”的情感分别是积极和消极的,整个句子因充斥着积极和消极两种情感而难以判断句子级情感. 而属性级情感分析则是分别判断针对这两个属性的情感,因此能够有效识别出更细粒度的情感信息.

属性级情感分类的关键在于识别与给定属性相关的上下文描述,并判断针对该属性的情感倾向. 然而,现实应用场景中许多评论对属性的情感表述是隐蔽的,并不包含明确的观点词 (opinion term), 故难以通过属性直接上下文推断出发文者针对该属性的情感极性. 这种不包含情感观点词,但仍能清楚地表达评论文本情感倾向的情感表达被称为隐式情感表达^[3]. 如表 1 所示的例子,“Great food, but the service is dreadful.”中的“great”和“dreadful”为情感观点词,通过该类词语可以轻易判断出属性的情感极性,故这类情感表达为显式情感表达;而评论“The two waitress’s looked like they had been sucking on lemons.”中,不包含针对“waitress”的显式观点词,人们需要理解整体上下文语义才能推断出发文者对“waitress”的情感是消极的,这类情感表达方式属于隐式情感表达. 根据评论中是否存在情感观点词, Li 等人^[4]将 SemEval-2014 数据集^[1]划分为显式情感表达数据 (explicit sentiment expression, ESE) 和隐式情感表达数据 (implicit sentiment expression, ISE). 据统计,隐式情感表达的句子在评论文本中普遍存在,在 Restaurant-14 和 Laptop-14 数据集^[1]中分别占 27.47% 和 30.19%. 然而,现有的方法尚未对隐式情感表达问题进行充分考虑,这直接影响了属性情感的识别效果.

表 1 显式/隐式情感表达示例

情感表达类型	例子
显式情感表达	Great food but the service is dreadful.
隐式情感表达	The two waitress’s looked like they had been sucking on lemons.

注: 下划线的词为待判断情感的属性词,带颜色的词为情感词

从直觉来看,识别隐式表达的情感更加依赖于对句子整体语义的理解. 然而,现有的方法大多采用注意力机制捕捉评论文本中与评价属性相关的上下文描述^[5-7]. 注意力机制是通过计算属性词与句子中每个词之间的相关性,从而赋予不同词之间互斥的权重来获取属性相关信息的. 这样的做法往往会对句子中的某个词 (或某几个词) 赋予较高的权重,而忽略了句子整体的信息. 因此,基于注意力机制的方法通常只能捕捉针对特定属性的显式情感,而难以挖掘隐含的情感. 此外,大多数的方法都将属性词和句子分别建模,导致属性词的表示脱离了上下文语境,这影响了属性情感识别的效果. 为此,本文从属性建模角度提出了改进策略,我们从全局上下文表示中抽取得到属性表示,并用该属性表示与句子全局做交叉注意力计算以获得属性感知的句子表示. 这样的做法克服了现有方法将属性词与句子孤立建模的缺点. 在此基础上,本文还进一步融合了属性局部和句子全局两部分信息,以加强句子整体语义在判断属性情感中的作用.

此外,隐式与显式情感表达的样本之间存在较大的差异,在实验中平等地处理这两类数据将会对模型的收敛过程产生负面影响,并直接损害整体情感识别的效果^[8]. 大量研究表明,如果训练数据以一种更有意义的顺序呈现,而不是随机排列,不仅适合人类的学习过程,而且更有利于机器的训练. 这种学习方式被称为课程学习 (curriculum learning, CL)^[9]. 课程学习概念最初是由 Bengio 等人^[9]提出的,其基本思想是让模型以由易到难的顺序学习. 在属性级情感分类问题中,识别显式与隐式表达的情感之间存在天然的难度差异. 因此,我们可以在模型训练过程中采用课程学习策略,按照由易到难的顺序学习这两类数据,从而达到提高情感分类性能的目的.

综上所述,为了解决隐式情感表达难以建模的问题,本文分别从属性建模方式和模型训练方法两个角度提出了改进方案:一种利用属性局部上下文与句子全局信息交叉建模,同时采用课程学习策略的属性级情感识别方法. 在 SemEval-2014^[1]和 MAMS^[10]两个数据集上的结果验证了本文方法的有效性. 总的来说,本文的主要贡献体现在

以下 3 个方面.

- 属性词和句子全局上下文内容被同时建模, 并利用交叉注意力机制获得融合属性信息的句子表示, 改进了现有方法将属性词和上下文孤立建模的缺点.

- 将课程学习方法用于解决隐式情感表达问题, 并提出了一种新的课程安排策略.

- 在 SemEval-2014 和 MAMS 两个数据集上的实验结果表明, 本文所提出的方法能够有效建模属性情感, 提升了情感分类的性能, 尤其是隐式情感表达的识别效果.

本文第 1 节介绍属性级情感分类和课程学习的相关工作和研究现状. 第 2 节介绍模型的具体结构和课程安排详细内容. 第 3 节介绍模型的实现细节以及在 SemEval-2014 和 MAMS 两个数据集上的实验结果. 第 4 节对模型进行进一步消融分析. 第 5 节总结全文.

1 相关工作

1.1 属性级情感分类相关工作

近年来, 国内外学者针对属性级情感分类做了大量工作. 研究方法从基于知识工程的方法, 到统计学习方法^[11,12], 再到深度学习方法^[13-16], 各种方法从不同的角度推动了该方向的快速发展. 与传统的情感分析任务不同, 属性级情感分析是针对特定的属性进行情感极性识别, 而不是判断整个句子的情感倾向. 其难点是如何识别和利用与给定属性相关的上下文描述, 以判断该属性的情感极性. 神经网络模型因其能以端到端的方式自动学习重要特征表示而成为属性级情感分类任务的主流解决方法. 许多工作利用注意力机制捕捉给定属性相关的上下文内容^[5-7,14,17]. Wang 等人^[5]将属性词嵌入与句子中每个词的词嵌入拼接, 并利用属性词嵌入表示计算注意力权重, 以获得给定属性的重要信息. Tang 等人^[6]提出了一种深度记忆网络模型, 将属性上下文内容和其位置信息考虑在内, 通过多层注意力机制叠加获取与给定属性相关的信息. Ma 等人^[7]引入一种交互注意力模型, 对句子上下文和属性词进行交互学习, 分别生成属性词和整体句子的表示. 这些方法的特点是先分别建模属性词和评论句子的特征表示, 然后利用注意力机制捕捉句子中与属性相关的信息, 进而判断属性级情感.

虽然早期各种基于注意机制的方法在属性情感分类中取得了不错的效果^[5-7,17,18], 但这些工作往往将属性词和句子整体分开建模, 使得属性词的表示脱离了上下文语境. 最近, Zeng 等人^[19]提出了一种局部上下文关注机制 (local context focus, LCF). 他们认为, 越靠近属性的词对于判断属性情感的作用越大, 因此将两个词之间的间隔词数定义为语义相对距离 (semantic relative distance, SRD), 并通过人为设定 SRD 的大小以限定识别属性情感的上下文关注范围. 然而, 硬性地局部上下文关注容易丢失与属性词间隔较远信息, 从而造成情感极性误判. Phan 等人^[20]曾在 LCF 基础上做出改进, 提出了基于语法的局部上下文关注机制 (local context focus on syntax, LCFS). 与 LCF 不同, LCFS 把依存句法树中两个词之间的最短路径作为句法相对距离 (syntactic relative distance, SRD). 近年来, 随着预训练语言模型在自然语言处理各个任务中展现出强大的表现力, 许多工作开始利用预训练语言模型学习到的知识来增强模型的表现力^[10,20-23].

受 LCF 和 LCFS 机制的启发, 本文利用注意力机制计算属性词与句子中每个词之间的相关性, 并用注意力结果加权的全局上下文表示作为属性感知的句子表示, 以获得整体句子语境中和目标属性最相关的上下文语义. 与以往基于注意力机制的方法不同, 本文方法的属性表示是从全局上下文表示中抽取得到的, 这样避免了属性建模脱离上下文语境的缺点. 除此之外, 在得到属性感知的句子表示基础上, 我们进一步融合了句子全局信息, 以满足判断隐式情感对整体句子语义的需求.

1.2 课程学习相关工作

本文的工作还在训练过程中采用了课程学习以提高模型的最终性能, 改善隐式情感表达句子的情感识别准确率. 受人类和动物学习过程的启发, 课程学习方法的基本假设是, 当训练数据不是随机呈现, 而是以一种更有意义的顺序组织起来时, 神经网络能够训练得更好. 例如, 先用简单的数据训练模型, 然后利用难度更大的数据进行训练, 最后让模型在完整的数据集上训练至收敛. 这种学习策略已经在各种自然语言处理及相关任务中取得了非常

有效的结果^[24],如机器翻译^[25,26]、文本分类^[27]等。

课程学习相关的许多工作都聚焦于如何挖掘困难样本和如何设计训练过程中样本的呈现顺序^[28]。在样本难度衡量问题中,一种简单而又常用的措施是通过直接观察目标任务难度和数据结构特征,人为预定义难度衡量规则^[29]。在自然语言处理领域,文本长度是一种常见的数据难度衡量标准^[30,31]。如 Tay 等人^[31]认为,问答例子的段落长度反应了其难度,太长或太短的段落都会增加问答的难度。除此之外,Wei 等人^[27]认为数据增强的样本中含有大量噪声,可将其视为困难样本,并提出了一种课程数据增强方法,该方法提升了少样本分类的性能。在情感分析任务中,评论文本的情感极性都是通过上下文语境推断的。在显式情感表达的句子中存在明确的情感词,如“good”“dreadful”等,这类句子往往较容易判断情感,而在隐式情感表达的句子中却不包含明确的情感词,这类样本的表达往往更加隐晦,情感推断难度大。因此,根据这一思路,我们认为,将隐式情感表达数据和显式情感表达数据分别预定义为困难样本和简单样本,是十分自然的选择。

近年来,研究人员在课程安排方面也做了大量工作。通常,预定义的训练课程安排可以分为离散式和连续式^[32]两种。所谓离散式是指在训练固定间隔的 epoch 数 (>1) 后,或收敛到当前的子数据集后,调整训练数据集^[33],直到训练数据包含整个训练集。连续式课程安排则是在训练过程中的每个 epoch 调整训练数据集^[26],最终在完整的训练数据集上完成训练,直到模型收敛。基于以上两种策略,Wei 等人^[27]提出了双阶段式课程数据增强方法和渐进式课程数据增强方法。课程学习训练标准的最后应使所有样本的权重与原始数据分布一致,这也启发了课程安排算法的设计。本文在上述工作的基础上,提出了一种根据训练阶段调整不同难度样本损失权重大小的课程安排方式。这样的做法避免了直接对数据集的重构改动,并且调整的方式更加灵活且连续,详细内容可见本文第 2.3 节。

除此之外,面对隐式情感表达问题,Li 等人^[4]提出了一种基于监督对比预训练的解决方案。他们收集并利用了大量领域内的评论数据对 BERT 模型进行预训练。预训练任务包括监督对比学习、属性词掩码预测和评论重建。通过预训练,他们得到相同情感类别情感表达的对齐表示,以此提高模型在属性情感分类任务中的表征能力。本文的工作主要聚焦于预训练模型微调阶段的设计,并未涉及预训练部分的工作。在未额外使用大量领域数据,仅有较小训练开销的情况下,本文建立的方法能够有效提升隐式情感表达数据的情感识别性能。

2 本文方法

根据前面的介绍,本文所提的方法主要聚焦于情感识别模型框架的改进和模型训练方式的优化两方面,下面就相关模型结构和课程学习训练细节分别予以介绍。

2.1 任务定义及模型整体结构框架

属性级情感分类任务定义如下:给定句子序列 $s = \{w_1, \dots, w_l, \dots, w_n\}$ 及属性序列 $A = \{w_i, \dots, w_{i+m}\}$, 属性序列是句子序列的子序列,任务目标是预测句子 s 关于属性 A 的情感极性 $y \in \{1, \dots, C\}$ 。其中 m, n 分别为属性序列和句子序列的长度, C 表示情感类别的数量。输入采用“[CLS] 句子 [SEP] 属性 [SEP]”的形式^[34],其中“[CLS]”和“[SEP]”是 BERT 预训练语言模型输入序列的特殊记号,输入序列以“[CLS]”开始、“[SEP]”结束。将句子序列和属性序列分别视为两个片段,在中间加入一个“[SEP]”符号以分隔。

模型由 3 部分组成:(1) 句子与属性词编码模块;(2) 属性与全局交叉注意力计算模块;(3) 属性局部与句子全局融合编码模块,最终编码得到面向属性的句子语义特征向量用于分类。模型整体结构如后文图 1 所示。下面详细阐述各模块的实现细节。

2.2 模型结构实现

2.2.1 句子与属性词编码

为了充分利用从大规模语料中学习到的文本特征,本文采用 BERT 预训练语言模型^[35]作为句子编码器。输入句子经过 BERT 编码后得到上下文语义相关的句子全局表示 H , 记第 i 个词 w_i 经 BERT 编码后的词向量表示为 h_i 。在句子表示的基础上对每一个词进行掩码操作,即将属于属性序列的词赋予权重为 1,句子中其余词赋予权重为 0,得到属性序列表示 H^{aspect} 。其中 $H \in \mathbb{R}^{n \times d_{\text{model}}}$, $H^{\text{aspect}} \in \mathbb{R}^{n \times d_{\text{model}}}$, d_{model} 为 BERT 的词向量维度,本文采用 BERT-base

模型, 故 d_{model} 为 768 维. 这样从句子全局表示中经掩码得到的属性词表示是上下文语义相关的.

$$H^{\text{aspect}} = \text{mask}(H) = \begin{cases} h_i, & w_i \in A \\ 0, & w_i \notin A \end{cases} \quad (1)$$

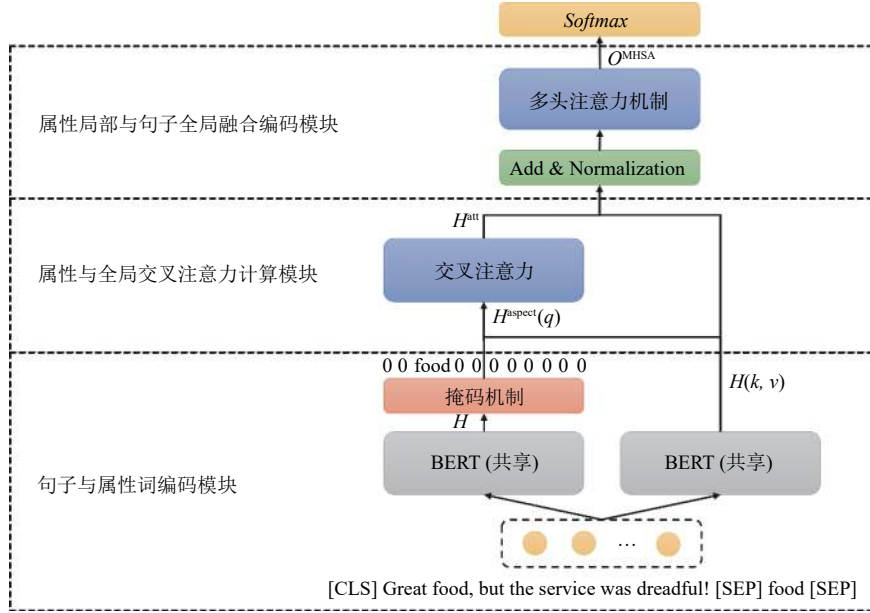


图 1 模型结构示意图

2.2.2 属性与句子全局交叉注意力计算

从全局上下文中提取与给定属性相关的内容有助于在多个属性的情况下区分针对不同属性的情感. 然而, 局部上下文关注机制^[19]通过控制语义相对距离的大小捕捉固定窗口内的内容, 而无法针对不同属性自适应地选择上下文关注内容, 容易丢失长距离的信息. 如图 2 所示的例子, 若给定属性为“the waiter”, 人为设定局部上下文关注窗口大小为 5, 则关注范围为“complained to the waiter and then”, 那么关键观点词“rudeness”将无法得到足够的关注. 这种类似的长程信息在评论文本中普遍存在. 因此, 本文采用注意力机制捕捉与评价属性相关的局部信息: 用属性词表示 H^{aspect} 作为查询向量 Q , 句子全局表示 H 作为键值向量 K, V 计算注意力权重, 进一步得到属性感知的句子表示 $H^{\text{att}} \in \mathbb{R}^{n \times d_{\text{model}}}$.

$$H^{\text{att}} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{\text{model}}}}\right) \cdot V = \text{Softmax}\left(\frac{H^{\text{aspect}} H^T}{\sqrt{d_{\text{model}}}}\right) \cdot H \quad (2)$$

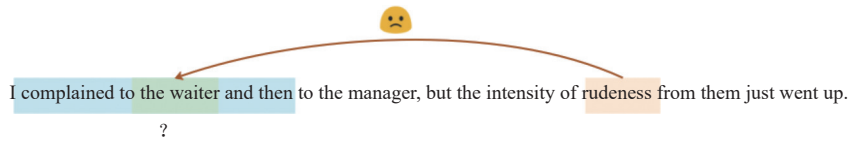


图 2 局部上下文关注窗口缺点示例

2.2.3 属性局部与句子全局融合编码

虽然属性级情感分析的关键在于抽取评论中与给定属性对象相关的内容, 但同时也不可缺少句子整体语义信息. 尤其是针对隐含情感表达, 判断这类句子的情感极性更加依赖于对句子整体语义的理解. 为此, 本文将注意力加权的属性感知句子表示 H^{att} 与句子全局表示 H 相加后进行多头自注意力机制^[36]编码, 以充分融合属性局部和句子全局信息. 与 Transformer 结构类似^[36], 在每一个子层后面都加入了层归一化^[37]处理.

$$O^{\text{MHSA}} = \text{MultiHead}\left(\text{LayerNorm}\left(H^{\text{att}} + H\right)\right) \quad (3)$$

2.3 模型训练优化

为了获得更好的性能, 高效的学习过程是必要的. 本文在预训练模型的微调阶段引入了课程学习策略, 并在优化目标中加入了三元组损失, 以实现更好的训练效果. 以下将详细介绍本文的课程安排方法.

2.3.1 样本损失重加权式课程安排

课程学习概念最初是由 Bengio 等人提出的^[9], 其原始定义如下: 课程是 T 个训练步骤上的一个训练标准序列 $C = (Q_1, \dots, Q_i, \dots, Q_T)$. 每个标准 Q_i 是对训练目标分布 $P(z)$ 的重新加权. 因此, $Q_i(z)$ 可以表示为:

$$Q_i(z) \propto W(z)P(z), \quad \forall z \in D \quad (4)$$

其中, z 表示训练数据集 D 中的一个样本, $W(z)$ 为权重函数.

课程学习的通用框架为“样本难度衡量+课程安排”. 如第 1.2 节所述, 由于隐式情感表达较显式情感表达更为隐晦委婉, 情感识别难度大, 我们将隐式情感表达数据视为困难样本, 显式情感表达数据作为简单样本. 课程安排采用样本损失重加权策略. 这种方式无需打乱数据集, 仅在训练的不同阶段对不同难易程度的训练样本给予不同大小的惩罚力度. 具体实现如下: 在训练开始阶段, 我们希望模型更加关注简单样本, 故赋予简单样本更大的损失权重; 随着训练地不断深入, 模型将提高对困难样本的关注, 因此增加困难样本的损失惩罚系数; 最后, 所有样本的损失权重大小将达到平衡, 即训练的最后阶段是在完整的训练数据集上完成的. 样本权重的设计可以有多种不同的方式, 可将其视为一种超参数. 本文采用最简单的线性方式, 简单样本和困难样本的损失权重变化函数分别为:

$$\lambda_{\text{easy}} = \max\left\{1, 2 - \frac{i}{N}\right\} \quad (5)$$

$$\beta_{\text{hard}} = \min\left\{1, e^{-4} + \frac{i}{N}\right\} \quad (6)$$

其中, i 表示第 i 个课程阶段, N 表示课程阶段总数, $i \in [0, N]$. 由以上公式可以实现简单样本的损失权重 λ_{easy} 从 2 线性降至 1, 困难样本的损失权重 β_{hard} 则从 e^{-4} 线性增至 1. 最终实现简单样本和困难样本的损失权重都为 1, 即在完整的训练数据集上进行训练.

2.3.2 优化目标

正如前文所述, 我们在优化目标中加入了三元组损失 (triplet loss). 三元组损失是一种在分类任务中常见的优化方法. 该方法首先在人脸识别问题中被提出^[38], 目前在 NLP 领域也得到了广泛应用^[39]. 其基本思路是, 三元组网络试图通过模型训练, 使基准样本与同类别样本在特征空间中的距离更近, 与不同类别样本之间的距离更远, 从而学习到更具有区分度的特征表示. 假设从数据集中任意构造一个三元组 (a, p, n) , 其中 a 为基准样本, p 为与 a 相同类别但不同的正样本, n 为与 a 不同类别的负样本, 我们希望:

$$d(a, p) + \text{margin} < d(a, n) \quad (7)$$

$$d(i, j) = \left\|f(x_i) - f(x_j)\right\|_2^2 \quad (8)$$

其中, $d(\cdot)$ 表示计算两个样本之间的距离, 本文采用欧式距离计算方式, 公式 (8) 中 $f(x_i)$ 和 $f(x_j)$ 分别表示样本 x_i, x_j 经网络编码后输出的最终特征表示; margin 是不同类别之间的边界阈值, 可以通过改变 margin 值控制正负样本的距离. 理论上较大的 margin 能够增强模型不同类别样本的区分度, 但过大的 margin 值会增加模型训练的难度, 需要在实验中确定 margin 的选择.

因此, 三元组损失的计算公式如下:

$$L^{\text{triplet}} = \sum_i^N \left[\left\|f(x_i^a) - f(x_i^p)\right\|_2^2 - \left\|f(x_i^a) - f(x_i^n)\right\|_2^2 + \text{margin}_+, \forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} \right] \quad (9)$$

其中, \mathcal{T} 为所有可能的三元组, 总共有 N 个三元组.

由定义可知, 若三元组已满足公式 (7), 则损失为 0, 可将其视为简单三元组. 这类三元组对训练没有贡献, 反

而影响了模型的收敛速度. 若负样本与基准样本间的距离大于正样本与基准样本间的距离, 但损失值还未达到 0, 即 $(d(a, p) < d(a, n) < d(a, p) + \text{margin})$, 则可通过训练不断降低损失值, 称这类三元组为一般三元组. 若负样本与基准样本间的距离小于正样本与基准样本之间的距离 $(d(a, n) < d(a, p))$, 此时损失值会出现较大的震荡, 这类三元组属于困难三元组^[38]. 通常, 一个数据集中可以构造非常多的三元组, 并且其中包含许多简单三元组. 在训练中需要选择合适的三元组, 否则会大大降低训练效率.

本文只选择困难三元组进行优化, \mathcal{T} 为所有可能的困难三元组. 另外, 单一使用三元组损失容易造成模型过拟合, 故总的损失函数包含交叉熵损失和三元组损失两部分. 最终的损失函数为:

$$L = \underbrace{\lambda_{\text{easy}} L_e^{\text{CE}} + \beta_{\text{hard}} L_h^{\text{CE}}}_{L^{\text{CE}}} + L^{\text{triplet}} \quad (10)$$

其中, λ_{easy} 和 β_{hard} 来自公式 (5) 和公式 (6), 分别为简单样本和困难样本的损失惩罚系数, L_e^{CE} 和 L_h^{CE} 分别为简单样本和困难样本的交叉熵损失.

3 实验分析

本节将对实验数据、模型实现细节以及模型在基准数据集上的实验结果进行详细介绍.

3.1 实验数据

本文在 3 个公开数据集上进行了验证. Laptop-14 和 Restaurant-14 数据集来自 SemEval-2014^[1], 是 ABSA 任务最常用的数据集. 本文用 Li 等人^[4]划分的 ESE 和 ISE 数据评估模型在显式和隐式情感表达数据上的表现. 由于 SemEval-2014 未公布验证数据, 我们从训练数据集中随机采样 20% 的样本作为验证集. 除此之外, MAMS (multi aspect multi sentiment) 数据集集中的每条评论均含有多个属性词, 且不同属性词情感极性不同, 是一个更具有挑战性的数据集^[10]. 本文采用该数据集分析模型在多属性、多情感场景下的性能. 以上 3 个数据集均包含用户评论文本、属性对象和属性对应的 4 类情感极性标签 (积极、消极、中性和矛盾). 与已有方法的处理方式一致, 因矛盾情感极性样本数量极少, 本文在实验中去除了该类情感标签的数据. 数据集详细信息如表 2 所示.

表 2 实验数据集

数据集		消极	中性	积极	总数	ISE占比 (%)
Restaurant-14	测试集	196	196	728	1120	23.84
	训练集	805	633	2164	3602	28.60
	总计	1001	829	2892	4722	27.47
Laptop-14	测试集	128	169	341	638	27.43
	训练集	866	460	987	2313	30.96
	总计	994	629	1328	2951	30.19
MAMS	训练集	3380	5024	2764	11168	—
	验证集	403	604	325	1332	—
	测试集	400	607	329	1336	—
	总计	4183	6235	3418	13836	—

3.2 超参数设置

为了验证本文方法的有效性, 模型中所有超参数均通过多次实验验证得到. BERT 模块使用 Huggingface Transformers 库中的预训练参数进行初始化, 选择 AdamW 优化器进行梯度更新, 学习率遵循文献 [35] 中的建议, 设定为 $2\text{E}-5$. 批量大小设置为 16, 最大序列长度为 85, dropout 率为 0.1, L2 正则系数为 $1\text{E}-3$, margin 大小设为 0.2. 课程学习阶段数目 N 默认设置为 5, 每一个阶段即为一个 epoch. 模型总训练 epoch 数为 20, 为防止模型过拟合, 在训练中还采取了提前停止策略, 若在 5 个 epoch 内验证集上的准确率没有提升则停止训练. 保留验证集上性能最好的模型在测试集上进行测试, 并报告测试集的结果.

3.3 基准模型

本文从不同角度将提出的模型与基线方法进行比较,以全面评估本文方法的性能.我们选取的基线模型如下.

- 基于注意力机制的方法: ATAE-LSTM^[5]、MemNet^[6]、RAM^[17]、IAN^[7].
- 基于预训练语言模型的方法: BERT-SPC^[35]、AEN-BERT^[34]、LCF(CDW/CDM)-BERT^[19].

为了更好地分析课程学习微调的效果,本文进一步对以下变体模型进行对比研究.

- CA-BERT: 本文提出的完整模型.
- CA-BERT+TL: 本文提出的模型,优化目标包含交叉熵损失和三元组损失.
- CA-BERT+CL: 本文提出的模型,优化目标包含交叉熵损失和三元组损失,并在微调中运用样本损失重加权式课程学习策略.

- BERT-SPC+TL: BERT-SPC 模型,优化目标包含交叉熵损失和三元组损失.
- BERT-SPC+CL: BERT-SPC 模型,优化目标包含交叉熵损失和三元组损失,训练采用样本损失重加权式课程学习策略.

3.4 实验结果

模型采用准确率 (accuracy) 和 $F1$ 值进行评估.为了减轻实验结果的偶然性,每一个模型均用 3 组随机种子进行训练,最终报告 3 组测试结果的平均值.本文在所有的基线模型上重新进行了实验,并报告了模型在 ISE 和 ESE 数据上的准确率.由于 MAMS 数据集未标注观点词,故无法区分显式和隐式情感表达.因此本文未对其进行课程学习研究,主要用该数据集分析模型在多属性、多情感场景下的表现力.基线模型和本文所提模型在 Restaurant-14、Laptop-14 和 MAMS 数据集上的性能如表 3 所示.

表 3 模型实验结果 (%)

模型	Restaurant-14			Laptop-14			MAMS	
	Acc/F1	ESE	ISE	Acc/F1	ESE	ISE	Acc/F1	
Attention	ATAE-LSTM ^[5]	75.65/63.54	82.26	54.68	67.71/60.78	73.22	53.14	63.67/62.08
	MemNet ^[6]	76.85/63.74	84.04	53.81	68.81/62.86	71.78	59.81	65.99/64.92
	RAM ^[17]	77.56/65.74	84.72	54.81	69.12/64.25	71.92	60.57	75.85/74.65
	IAN ^[7]	75.92/62.68	83.20	51.81	67.24/63.72	72.43	57.90	68.06/66.23
KnowledgeEnhanced	AEN-BERT ^[34]	81.87/71.56	90.07	57.05	76.33/71.03	80.99	65.14	71.08/70.39
	LCF(CDM)-BERT ^[19]	83.24/74.38	91.05	58.30	76.86/71.70	81.71	64.00	83.10/82.49
	LCF(CDW)-BERT ^[19]	83.78/75.54	91.01	60.67	77.38/72.33	82.43	64.00	82.54/81.96
	BERT-SPC ^[35]	83.75/74.79	91.72	58.30	76.91/71.71	81.86	63.81	81.36/80.77
Ours (本文方法)	+TL	84.41/75.60	91.44	61.92	77.17/72.09	81.72	65.15	82.23/81.58
	+CL	85.09/77.19	92.26	62.17	78.01/73.52	81.14	69.72	—/—
	CA-BERT	84.64/76.72	91.72	62.05	78.27/74.01	81.71	69.14	84.01/83.55
	+TL	85.03/77.19	91.95	62.92	78.11/73.50	82.00	67.81	84.36/83.78
	+CL	85.09/77.59	91.99	63.05	78.63/74.51	81.07	72.19	—/—

注: 1. “ISE”和“ESE”分别表示隐式情感表达数据和显式情感表达数据; 2. 因LCF-BERT原文献^[19]用测试集作为验证集, 论文报告的结果为验证集上的结果, 故数值偏高, 与本文所报告的测试集上的结果之间存在较大差距; 3. 最好的结果用黑色加粗字体标出

由实验结果可知:

- (1) 本文提出的方法在两个基准数据集上的结果均优于基线模型.

本文提出的方法都以相当大的优势优于基线模型.相较于性能最好的基线模型 LCF(CDW)-BERT, CA-BERT+CL 在 Restaurant-14 和 Laptop-14 两个基准数据集上分别提高了 1.31% / 2.05% (Acc/F1) 和 1.25% / 2.18% (Acc/F1). 并且, CA-BERT+CL 在 ISE 数据上的表现也普遍优于基线模型. 结果充分验证了本文所提方法的有效性.

(2) 在未引入额外参数和数据的情况下, 课程学习策略有效提高了模型的情感识别性能, 尤其显著改善了隐式情感表达数据的分类准确率.

与未采用课程学习的 BERT-SPC 和 CA-BERT 相比, BERT-SPC+CL 和 CA-BERT+CL 的性能得到了明显提升. 在 Restaurant-14 和 Laptop-14 上, BERT-SPC+CL 的准确率相较于 BERT-SPC 分别提高了 1.34% 和 1.1%, CA-BERT+CL 相比 CA-BERT 则分别提高了 0.45% 和 0.36%. 且未引入额外参数的 BERT-SPC+CL 方法, 实现了较 CA-BERT 更好的性能表现. 这充分验证了课程学习在提升模型分类性能中的有效性. 值得注意的是, 课程学习虽对 ESE 数据的影响甚微, 但在 ISE 数据上效果格外明显, 准确率至少提高了 1%. 这表明课程学习可以提高难样本的识别准确率.

除此之外, BERT-SPC+TL 和 CA-BERT+TL 在 3 个数据集中都有效提升了模型的分类性能, 这可以验证三元组损失在分类任务中发挥了重要作用. 但与仅加入三元组损失的效果相比, 结合课程学习方法的结果更优, 这说明本文方法性能的提升不仅得益于三元组损失的作用, 本文的课程安排策略依然起着重要作用.

(3) 属性与全局上下文的交叉注意力机制在多属性多情感场景下的性能表现更好.

CA-BERT 模型在 MAMS 数据集上达到了最优的性能, 相比 CDM-BERT 和 CDW-BERT 分别提高了 0.91%/1.06% (Acc/F1) 和 1.47%/1.59% (Acc/F1). 实验结果表明, 交叉注意力机制比局部上下文关注机制更能有效地提取不同属性的相关上下文. 事实上, 局部上下文关注只能捕捉固定窗口大小内的内容, 然而, 当评论中的属性个数增加时, 属性之间描述更加错综复杂, 局部上下文关注窗口不足以涵盖判断属性情感的足够信息. 而注意力机制根据句子中每个词与属性之间的相关性赋予相应权重, 这样更有利于捕捉长距离的内容. 因此, 属性词与全局上下文的交叉注意力机制在面对多属性多情感场景时更具优势. 另外, 其他基于注意力机制的方法都将属性词建模独立于句子整体语义的建模. CA-BERT 模型则是从句子全局表示中抽取得到属性词表示, 在性能上明显优于以上几个模型, 这也验证了本文模型在多属性多情感场景下的有效性.

4 模型消融分析

本节将进一步深入分析本文提出的方法, 主要包括以下几个有趣的问题: (1) 属性局部与句子全局融合编码模型是否起作用? 属性情感分析中属性局部信息与句子全局信息哪个更重要? (2) 课程学习阶段数 N 将如何影响模型的最终性能? 第 4.1 节对应问题 (1), 问题 (2) 将在第 4.2 节中进行阐述.

4.1 属性局部与句子全局融合编码消融分析

为了分析属性局部与句子全局两部分信息对属性情感分类及隐式情感表达识别性能的作用, 本文对 CA-BERT 模型中属性局部与句子全局融合编码模块进行消融分析. 我们分别去除句子全局信息和属性局部信息, 得到 CA-BERT (w/o Global) 和 CA-BERT (w/o Local) 两个变体模型, 并在 Restaurant-14 和 Laptop-14 两个基准数据集和 MAMS 数据集上进行实验. 实验结果见表 4.

表 4 模型消融实验结果 (%)

模型	Restaurant-14				Laptop-14				MAMS	
	Acc	F1	ESE (Acc)	ISE (Acc)	Acc	F1	ESE (Acc)	ISE (Acc)	Acc	F1
CA-BERT	84.64	76.72	91.72	62.05	78.27	74.01	81.71	69.14	84.01	83.55
w/o Global	80.39	67.98	89.53	51.19	76.07	70.59	80.85	63.43	82.29	81.89
w/o Local	84.40	76.13	91.72	61.05	77.27	72.18	82.44	63.62	83.76	83.24

注: 1. 为减轻实验偶然性, 所有实验结果均为 3 个不同随机种子结果的平均值; 2. w/o Global 和 w/o Local 分别表示没有句子全局信息和没有属性局部信息; 3. ESE 和 ISE 数据上仅报告了准确率

根据表 4 的实验结果, 我们可以得出以下重要信息.

(1) 属性局部与句子全局两部分信息对属性情感分类都至关重要, 任一方面信息的缺失都会对模型的情感识别性能产生负面影响, 其中句子全局信息更为重要.

CA-BERT (w/o Global) 在 Restaurant-14 数据上性能下降了 4.25% / 8.74% (Acc/F1), 而 CA-BERT (w/o Local)

仅下降了 0.24% /0.59% (Acc/F1); 同样, 在 Laptop-14 数据集上, CA-BERT(w/o Global) 性能下降了 2.20% /3.42% (Acc/F1), 而 CA-BERT(w/o Local) 仅下降了 1.00% /1.83% (Acc/F1). 融合属性局部和句子全局信息在 3 个数据集上都实现了最佳的性能. 对比以上实验结果可知, 句子全局信息的缺失对分类的最终性能影响更大. 而现有的许多方法都聚焦于如何有效提取属性局部信息, 忽略了对句子整体语义的充分利用.

(2) 识别隐含表达的情感更加依赖于句子全局语义信息.

CA-BERT (w/o Global) 在 Restaurant-14 ISE 和 Laptop-14 ISE 数据上的准确率下降分别达到了 10.86% 和 5.71%, 而 CA-BERT(w/o Local) 仅分别下降了 1.00% 和 5.52%. 这与我们的直觉是相符的, 在判断隐晦的情感时, 人类也需要借助更多的句子整体信息.

(3) 多属性多情感场景下, 句子全局信息依然占据更大的作用.

值得注意的是, 在多属性多情感场景下, 句子全局信息依然占据更大的作用. 在 MAMS 数据集上, 属性局部信息的缺失使得模型的准确率下降了 0.25%, 而句子全局信息的缺失却使模型准确率下降了 1.72%. 因此, 尽管属性级情感分析是判断针对给定属性的情感, 但我们依然不可忽视对整体句子语义的理解和把握.

4.2 课程学习阶段数影响分析

本文所提出的课程学习方法需要预先定义课程阶段数目 N , 它决定了训练的每个 epoch 中样本的损失权重大小. N 越大, 课程学习的划分粒度也越细. 本文选择以 BERT-SPC 模型为对象, 以 Restaurant-14 和 Laptop-14 数据集为例进行实验, 研究参数 N 对模型性能的影响. 我们在课程学习阶段数目 N 从 0 到 7 的变化范围内进行研究, 其中 $N=0$ 代表没有使用课程学习. 实验结果如图 3 所示.

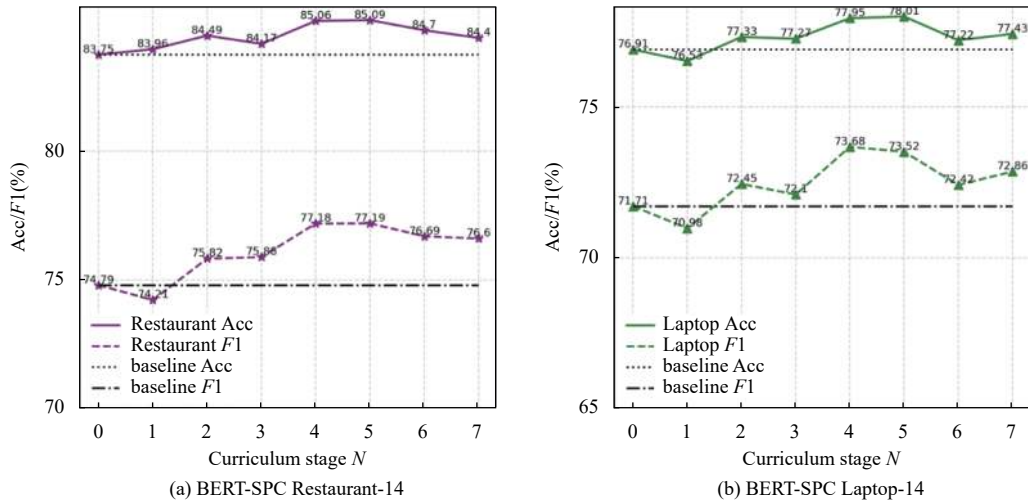


图 3 基于 BERT-SPC+CL 模型的课程学习阶段数目 N 影响分析图

当课程学习安排阶段数较小时, 课程学习不利于模型的训练. 如 $N=1$, 采用课程学习的训练方式反而低于基线模型 ($N=0$). 其原因是阶段数目小, 不同难度的数据在课程的不同阶段会发生损失权重大小的突变, 模型训练会产生较大的波动, 进而影响了模型最终的性能表现.

当课程学习安排阶段数增大时, 每一个课程阶段的样本损失权重差异变小, 不同课程阶段之间过渡更加平缓. 但课程阶段数目过大时, 不同课程阶段的差异过小, 课程学习优势不凸显, 模型的性能也相对有所下降. 综合实验分析可知, 课程安排阶段数 $N=5$ 为本模型的最佳选择.

5 结束语

本文针对属性级情感分类任务中属性建模缺陷和隐式情感表达这两个问题进行了研究, 得到如下重要结论.

(1) 针对现有方法将属性与句子分开建模的缺点, 本文在局部上下文关注机制基础上进行了优化, 利用属性与全局交叉注意力机制捕捉属性相关局部上下文内容, 并将属性局部信息与句子全局信息相互融合进行文本建模. 实验结果证明了本文所提方法的有效性. 消融实验分析表明, 属性级情感分类需要属性局部上下文和句子全局上下文两部分的信息, 任一方面信息的缺失都将影响模型最终的分类效果. 虽然属性局部上下文信息在面对多属性情况下起着重要作用, 但尤其不可忽视句子全局的信息在属性情感分类中的关键作用.

(2) 针对隐式情感表达在评论中广泛存在, 且分类难度的问题, 本文探索了课程学习方法对于解决该问题的有效性, 并提出了一种样本损失重加权式的课程安排策略. 实验表明, 本文所提出的方法能够有效提升隐式情感表达数据的分类性能.

在下一步工作中, 我们将进一步探索在模型预训练阶段如何运用课程学习方法改善隐式情感表达的识别性能, 并研究如何自动衡量样本的难度以及自动课程安排的方法, 而非人为预定义样本难度衡量规则和课程.

References:

- [1] Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S. SemEval-2014 task 4: Aspect based sentiment analysis. In: Proc. of the 8th Int'l Workshop on Semantic Evaluation (SemEval 2014). Dublin: Association for Computational Linguistics, 2014. 27–35. [doi: [10.3115/v1/S14-2004](https://doi.org/10.3115/v1/S14-2004)]
- [2] Zong CQ, Xia R, Zhang JJ. Text Data Mining. 2nd ed., Beijing: Tsinghua Press, 2022. 150–159 (in Chinese).
- [3] Russo I, Caselli T, Strapparava C. SemEval-2015 task 9: CLIPeval implicit polarity of events. In: Proc. of the 9th Int'l Workshop on Semantic Evaluation (SemEval 2015). Denver: Association for Computational Linguistics, 2015. 443–450. [doi: [10.18653/v1/S15-2077](https://doi.org/10.18653/v1/S15-2077)]
- [4] Li ZY, Zou YC, Zhang C, Zhang Q, Wei ZY. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training. In: Proc. of the 2021 Conf. on Empirical Methods in Natural Language Processing. Punta Cana: Association for Computational Linguistics, 2021. 246–256. [doi: [10.18653/v1/2021.emnlp-main.22](https://doi.org/10.18653/v1/2021.emnlp-main.22)]
- [5] Wang YQ, Huang ML, Zhu XY, Zhao L. Attention-based LSTM for aspect-level sentiment classification. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 606–615. [doi: [10.18653/v1/D16-1058](https://doi.org/10.18653/v1/D16-1058)]
- [6] Tang DY, Qin B, Liu T. Aspect level sentiment classification with deep memory network. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 214–224. [doi: [10.18653/v1/D16-1021](https://doi.org/10.18653/v1/D16-1021)]
- [7] Ma DH, Li SJ, Zhang XD, Wang H. Interactive attention networks for aspect-level sentiment classification. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence, IJCAI-17. Melbourne: IJCAI.org, 2017. 4068–4074. [doi: [10.24963/ijcai.2017/568](https://doi.org/10.24963/ijcai.2017/568)]
- [8] Fadaee M, Monz C. Back-translation sampling by targeting difficult words in neural machine translation. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 436–446. [doi: [10.18653/v1/D18-1040](https://doi.org/10.18653/v1/D18-1040)]
- [9] Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In: Proc. of the 26th Annual Int'l Conf. on Machine Learning. Montreal: ACM Press, 2009. 41–48. [doi: [10.1145/1553374.1553380](https://doi.org/10.1145/1553374.1553380)]
- [10] Jiang QN, Chen L, Xu RF, Ao X, Yang M. A challenge dataset and effective models for aspect-based sentiment analysis. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 6280–6285. [doi: [10.18653/v1/D19-1654](https://doi.org/10.18653/v1/D19-1654)]
- [11] Kiritchenko S, Zhu XD, Cherry C, Mohammad S. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In: Proc. of the 8th Int'l Workshop on Semantic Evaluation (SemEval 2014). Dublin: Association for Computational Linguistics, 2014. 437–442. [doi: [10.3115/v1/S14-2076](https://doi.org/10.3115/v1/S14-2076)]
- [12] Vo DT, Zhang Y. Target-dependent Twitter sentiment classification with rich automatic features. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 1347–1353.
- [13] Tang DY, Qin B, Feng XC, Liu T. Effective LSTMs for target-dependent sentiment classification. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 3298–3307.
- [14] Wang JJ, Li J, Li SS, Kang YY, Zhang M, Si L, Zhou GD. Aspect sentiment classification with both word-level and clause-level attention networks. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: AAAI Press, 2018. 4439–4445.
- [15] Yu JF, Jiang J, Xia R. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. IEEE/ACM Trans. on Audio, Speech, and Language Processing, 2019, 27(1): 168–177. [doi: [10.1109/TASLP.2018.2875170](https://doi.org/10.1109/TASLP.2018.2875170)]

- [16] Cai HJ, Tu YF, Zhou XS, Yu JF, Xia R. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 833–843. [doi: [10.18653/v1/2020.coling-main.72](https://doi.org/10.18653/v1/2020.coling-main.72)]
- [17] Chen P, Sun ZQ, Bing LD, Yang W. Recurrent attention network on memory for aspect sentiment analysis. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 452–461. [doi: [10.18653/v1/D17-1047](https://doi.org/10.18653/v1/D17-1047)]
- [18] Feng C, Li HH, Zhao HY, Xue Y, Tang JY. Aspect-level sentiment analysis based on hierarchical attention and gate networks. Journal of Chinese Information Processing, 2021, 35(10): 128–136 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2021.10.015](https://doi.org/10.3969/j.issn.1003-0077.2021.10.015)]
- [19] Zeng BQ, Yang H, Xu RY, Zhou W, Han XL. LCF: A local context focus mechanism for aspect-based sentiment classification. Applied Sciences, 2019, 9(16): 3389. [doi: [10.3390/app9163389](https://doi.org/10.3390/app9163389)]
- [20] Phan MH, Ogunbona PO. Modelling context and syntactical features for aspect-based sentiment analysis. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 3211–3220. [doi: [10.18653/v1/2020.acl-main.293](https://doi.org/10.18653/v1/2020.acl-main.293)]
- [21] Xu H, Liu B, Shu L, Yu P. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 2324–2335. [doi: [10.18653/v1/N19-1242](https://doi.org/10.18653/v1/N19-1242)]
- [22] Yan H, Dai JQ, Ji T, Qiu XP, Zhang Z. A unified generative framework for aspect-based sentiment analysis. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Association for Computational Linguistics, 2021. 2416–2429. [doi: [10.18653/v1/2021.acl-long.188](https://doi.org/10.18653/v1/2021.acl-long.188)]
- [23] Sun C, Huang LY, Qiu XP. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 380–385. [doi: [10.18653/v1/N19-1035](https://doi.org/10.18653/v1/N19-1035)]
- [24] Xu BF, Zhang LC, Mao ZD, Wang Q, Xie HT, Zhang YD. Curriculum learning for natural language understanding. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 6095–6104. [doi: [10.18653/v1/2020.acl-main.542](https://doi.org/10.18653/v1/2020.acl-main.542)]
- [25] Lu JL, Zhang JJ. Exploiting curriculum learning in unsupervised neural machine translation. In: Proc. of the 2021 Association for Computational Linguistics. Punta Cana: Association for Computational Linguistics, 2021. 924–934. [doi: [10.18653/v1/2021.findings-emnlp.79](https://doi.org/10.18653/v1/2021.findings-emnlp.79)]
- [26] Platanios EA, Stretcu O, Neubig G, Poczos B, Mitchell T. Competence-based curriculum learning for neural machine translation. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 1162–1172. [doi: [10.18653/v1/N19-1119](https://doi.org/10.18653/v1/N19-1119)]
- [27] Wei J, Huang CY, Vosoughi S, Cheng Y, Xu SQ. Few-shot text classification with triplet networks, data augmentation, and curriculum learning. In: Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021. 5493–5500. [doi: [10.18653/v1/2021.naacl-main.434](https://doi.org/10.18653/v1/2021.naacl-main.434)]
- [28] Koemi T, Bojar O. Curriculum learning and minibatch bucketing in neural machine translation. In: Proc. of the 2017 Int'l Conf. Recent Advances in Natural Language Processing. Varna: INCOMA Ltd., 2017. 379–386. [doi: [10.26615/978-954-452-049-6_050](https://doi.org/10.26615/978-954-452-049-6_050)]
- [29] Hachohen G, Weinshall D. On the power of curriculum learning in training deep networks. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 2535–2544.
- [30] Spitkovsky VI, Alshawi H, Jurafsky D. From baby steps to leapfrog: How “less is more” in unsupervised dependency parsing. In: Proc. of the Human Language Technologies: The 2010 Annual Conf. of the North American Chapter of the Association for Computational Linguistics. Los Angeles: Association for Computational Linguistics, 2010. 751–759.
- [31] Tay Y, Wang SH, Luu AT, Fu J, Phan MC, Yuan XD, Rao JF, Hui SC, Zhang A. Simple and effective curriculum pointer-generator networks for reading comprehension over long narratives. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 4922–4931. [doi: [10.18653/v1/P19-1486](https://doi.org/10.18653/v1/P19-1486)]
- [32] Wang X, Chen YD, Zhu WW. A survey on curriculum learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(9): 4555–4576. [doi: [10.1109/TPAMI.2021.3069908](https://doi.org/10.1109/TPAMI.2021.3069908)]
- [33] Shu Y, Cao ZJ, Long MS, Wang JM. Transferable curriculum for weakly-supervised domain adaptation. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI Press, 2019. 4951–4958. [doi: [10.1609/aaai.v33i01.33014951](https://doi.org/10.1609/aaai.v33i01.33014951)]

- [34] Song YW, Wang JH, Jiang T, Liu ZY, Rao YH. Targeted sentiment classification with attentional encoder network. In: Proc. of the 28th Int'l Conf. on Artificial Neural Networks. Munich: Springer, 2019. 93–103. [doi: [10.1007/978-3-030-30490-4_9](https://doi.org/10.1007/978-3-030-30490-4_9)]
- [35] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- [36] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- [37] Ba JL, Kiros JR, Hinton GE. Layer normalization. arXiv:1607.06450, 2016.
- [38] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 815–823. [doi: [10.1109/CVPR.2015.7298682](https://doi.org/10.1109/CVPR.2015.7298682)]
- [39] Ein Dor L, Mass Y, Halfon A, Venezian E, Shnayderman I, Aharonov R, Slonim N. Learning thematic similarity metric from article sections using triplet networks. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 2: Short Papers). Melbourne: Association for Computational Linguistics, 2018. 49–54. [doi: [10.18653/v1/P18-2009](https://doi.org/10.18653/v1/P18-2009)]

附中文参考文献:

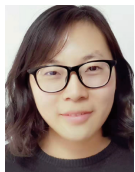
- [2] 宗成庆, 夏睿, 张家俊. 文本数据挖掘. 第2版, 北京: 清华大学出版社, 2022. 150–159.
- [18] 冯超, 黎海辉, 赵洪雅, 薛云, 唐婧尧. 基于层次注意力机制和门机制的属性级情感分析. 中文信息学报, 2021, 35(10): 128–136. [doi: [10.3969/j.issn.1003-0077.2021.10.015](https://doi.org/10.3969/j.issn.1003-0077.2021.10.015)]



叶静(2000—), 女, 硕士, 主要研究领域为自然语言处理, 情感分析.



宗成庆(1963—), 男, 博士, 研究员, 博士生导师, CCF 会士, 主要研究领域为自然语言处理, 机器翻译, 情感分析.



向露(1988—), 女, 博士, 助理研究员, 主要研究领域为自然语言处理, 文本生成, 人机对话系统.