

文章编号: 1003-0077(2024)03-0163-10

基于动态常识推理与多维语义特征的幽默识别

吐妮可·吐尔逊¹, 闵昶荣¹, 林鸿飞¹, 张冬瑜², 杨亮¹

(1. 大连理工大学 计算机科学与技术学院, 辽宁 大连 116024;

2. 大连理工大学 软件学院, 辽宁 大连 116620)

摘要: 随着社交媒体的飞速发展, 幽默识别任务受到研究者广泛关注。其目标是判断给定文本是否具有幽默表达。现有方法主要基于幽默产生理论, 采用规则或神经网络模型提取多种幽默相关特征, 如不一致性、情感和语音等。然而, 这些方法未充分捕捉文本内部的情感特征, 忽视了隐含在幽默文本中的情感表达, 影响了幽默识别的准确性。为解决此问题, 该文提出了 CMSOR 方法, 以动态常识和多维语义特征为驱动。首先, 利用外部常识信息从文本中动态推理说话者的隐式情感表达; 然后, 引入 WordNet 词典计算词级语义距离, 捕捉不一致性, 并计算模糊性特征; 最后, 基于这三个特征维度构建幽默语义, 实现幽默识别。实验证明, CMSOR 模型相对于当前基准模型在三个公开数据集上的识别性能均有显著提升。

关键词: 幽默识别; 常识推理; 模糊理论; 注意力机制

中图分类号: TP391

文献标识码: A

Humor Recognition Based on Dynamic Commonsense Reasoning and Multi-Dimension Semantic Features

Tuerxun · TUNIKE¹, MIN Changrong¹, LIN Hongfei¹, ZHANG Dongyu², YANG Liang¹

(1. School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China;

2. School of Software Technology, Dalian University of Technology, Dalian, Liaoning 116620, China)

Abstract: As an emerging topic in NLP, humor recognition is to discriminate whether a given text expresses humor. To fully capture emotional features within the text, we propose CMSOR method based on dynamic commonsense reasoning and multi-dimension semantic features. It adopts the commonsense to infer latent emotion feature of speakers from the text, then leverages WordNet lexicon to calculate word level distances as the inconsistent features and the ambiguous features. We make use of these three humor-specific features to construct humor semantics. Experiments on three publicly available benchmarks demonstrate that CMSOR is superior to state-of-the-art models.

Keywords: humor recognition; commonsense reasoning; ambiguity theory; attention mechanism

0 引言

幽默作为一种修辞手法, 是人类交际中不可或缺的一部分, 在使得人与人之间的沟通更加流畅的同时, 营造了轻松愉悦的交流氛围。得益于社交媒体飞速发展所带来的海量文本数据, 自然语言处理领域的文本幽默识别研究在近年来取得了长足进展。文本幽默识别的主要目标是通过计算方法来理解文本中的幽默表达并判断该文本是否为幽默。幽

默识别不仅能够应用于文本生成、机器翻译以及隐喻识别等任务, 还能够赋予机器理解幽默的能力, 提升现实中人机交互的效果。因此, 从文本中理解幽默产生的机制并识别幽默文本变得尤为重要。

从语言学与心理学的角度, 主要存在三种观点来解释幽默的产生, 分别是: 优越论^[1]、宽慰论^[2]以及乖讹论^[3]。其中, 优越论认为幽默是一种表达并强调自我价值与地位的方式, 它强调通过取笑、讽刺或嘲笑他人来获取优越感; 宽慰论认为幽默有助于缓解人们的压力和紧张情绪; 乖讹论又称为不一致

收稿日期: 2023-12-01 定稿日期: 2023-12-07

基金项目: 国家自然科学基金(62076046, 62076051, 62376051)

性理论,它的表达中通常会包含一些出人意料的 inconsistency,通过产生违背人们常识和期望的事物的感知,来引发人们的笑声和关注。基于上述理论,研究者们从多个角度提取文本中的幽默特征,同时通过设计不同结构的神经网络模型来学习幽默的深层次语义,基于此判断该文本是否为幽默。例如,Chauhan 等人^[4]认为幽默与情感和情绪密切相关,提出了利用 Transformer 和情绪感知嵌入(SE-Embedding)的多任务框架来进行幽默检测。Liu 等人^[5]基于“优越论”和“宽慰论”的观点,结合情感特征对语篇单元中的情感关系建模,证明了情感信息能更有效地解决对话幽默识别问题。Li 等人^[6]使用“乐观幽默类型”和“悲观幽默类型”的情感极性来标注数据集中“积极”和“消极”情绪类别,采用双向长短时记忆网络结合注意力网络的方法,捕捉俚语和微博表情符号在情感分析中的影响,为深入了解俚语和微博表情符号对中文情感分析提供了新视角。

从上述工作中可知,文本内蕴含的情感特征对于识别幽默表达十分重要,这些工作主要通过外部词典匹配的方式来捕捉文本内的情感特征。然而,本文发现在幽默表达中很多情绪往往是隐式表达的,如表 1 所示,其中第二个幽默样本表达了“悲伤”或者“愤怒”的情绪,但是该样本并没有包含直接表达情绪的词汇,而是通过短语“got fired”来表达,这种方式称为隐式情感表达。现存的幽默识别方法主要采用外部情感词典来捕捉文本内的情感信息。显然,这种方式无法有效识别出这些隐式情感表达,降低了模型识别文本幽默的能力。

表 1 幽默样本以及包含的情绪信息

幽默文本	情绪信息
I used to play piano by ear, but now I use my hands.	自嘲
I can't believe I got fired from the calendar factory. All I did was take a day off !	愤怒
I don't trust people who do acupuncture. They're back stabbers.	不满

从认知角度,理解这些隐式情绪表达不仅需要结合上下文信息,还需要充分利用外部常识。尽管现有的预训练语言模型(PLM)能够高效地捕捉文本的上下文信息,但是由于其是在大规模通用语料上训练,因此无法有效感知这些文本背后的隐式情绪。

为了解决这一问题,本文提出一种动态常识与

多维语义特征驱动的幽默识别方法(Commonsense and Multi-dimensional Semantics Based Humor Detector, CMSOR)。该方法主要是利用外部常识,根据文本的上下文信息,动态地推断文本中的隐式情绪,并将其作为文本情绪特征的一部分,参与幽默识别。

具体地,该方法首先根据文本内容利用预训练常识推理工具 COMET^[7]根据上下文信息动态推断文本的内蕴情感信息,然后将文本内容与推断出的情感信息拼接融合,通过预训练语言模型 BERT 进一步将显式情感融入到文本语义当中,形成显式情感增强的文本表示。同时,利用外部词典 WordNet^[8]计算语义距离以及同义词数量,分别形成文本的不一致性特征以及模糊性特征。最后,将上述三种特征进行结合,形成多维幽默语义表示,输入到分类器中,得到幽默预测结果。

本文研究工作主要内容如下:

(1) 提出了一种动态常识驱动的幽默识别方法 CMSOR,利用外部常识动态捕捉文本的隐式情感特征,同时利用外部词典建模模糊性与不一致性特征,从多个维度构建幽默语义,实现幽默识别。

(2) 在 Pun of the Day、SemEval21 以及 CoBERT 三个公开数据集上进行了实验,实验结果表明,本文所提出的 CMSOR 模型相比于现有方法在四项评价指标上有明显提升,证明了方法的有效性。

1 相关工作

由于幽默表达本身的复杂性,幽默识别在近些年来一直是一项极具挑战性的任务。早期的幽默识别方法主要是基于特征工程,将统计机器学习方法作为分类器,在幽默理论的基础上设计不同的幽默特征提取方案。这些人工提取的特征包括通用语言学特征以及面向幽默的文本特征。例如, Mihalcea 和 Strapparova^[9]定义了头韵、反义词和成人俚语三种幽默特征,通过实验证明了他们在 one-liner 数据集中幽默识别的有效性。Mihalcea 等人^[10]将幽默文本分为“铺垫”和“笑点”两部分,通过计算两者的语义相关性进行幽默识别。Yang 等人^[11]深入探讨幽默潜在语义特征,构造了四种幽默特征,分别是语音特征、歧义特征、不一致性特征和情感特征。Morales 和 Zhai^[12]针对 Yelp 评论使用概率模型结

合背景文本资源进行幽默识别。Cattle 和 Ma^[13] 利用单词关联的语义关联特征进行幽默识别。上述这些工作大多是利用统计或者匹配的方法来提取文本中的浅层幽默特征,无法对幽默的深层次潜在语义进行表示,从而限制了幽默识别的性能。

随着计算能力的提高以及社交媒体数据的增长,深度学习在不同领域被广泛用于辅助或替代传统的特征工程。与其他领域相比,深度学习在幽默识别任务中应用较晚。这些基于深度神经网络的幽默识别方法主要是利用预训练语言模型表示文本,然后设计不同结构的神经网络实现对于幽默特征的深层次提取。例如,Bertero 等人^[14]认为幽默情景剧是一种具有独特特点的喜剧形式,背景笑声可以视为观众对于搞笑场景的反应,自动标注这些笑声可以有效地识别笑点,便在此基础上使用长短时记忆网络(LSTM)对幽默情景剧中的对话进行建模,同时提取对话语义特征和声音特征来识别笑点。Cuza 和 Buenod 等人^[15]针对西班牙推文结合了语言特征和基于注意力的递归神经网络进行幽默识别。Blinov 等人^[16]收集大量笑话和趣味对话构造俄语数据集,并微调语言模型用于幽默识别。Kao 和 Levy 等人^[17]提出模糊性和独特性两个特征,使用语言模型识别幽默语句。Weller 和 Seppi^[18]使用 Transformer 架构识别幽默。Hasan 等人^[19]使用循环神经网络进行多模态幽默识别。Diao 等人^[20]提出一种基于不一致性、模糊性、情感因素和语言学潜在语义结构的识别模型。Fan 等人^[21]基于 Bi-GRU 网络融合语音特征和歧义性特征进行幽默检测。Annamoradnejad 和 Zoghi^[22]改进 BERT 模型在自创建的幽默数据集 ColBERT 上进行实验,证实了提出的模型能够有效地检测幽默。Zhang 等人^[23]利用卷积神经网络结合标签转移关系提出多任务学习模型识别幽默。Ren 等人^[24]结合幽默和双关语识别任务,提出一种基于注意力的多任务学习模型来进行幽默检测。Ren 等人^[25]提出一种基于注意力机制的神经网络来验证发音、句法与词法特征对于幽默识别任务的重要性。

与上述工作类似,本文同样考虑了情感特征在幽默表达中的重要作用。不同的是,本文针对幽默文本中隐式情感表达难以被词典有效识别的问题,采用动态常识推理,从文本中推断内蕴的隐式情感,并结合模糊特征与不一致特征,从多个维度对于文本的幽默语义进行刻画。

2 动态常识与多维语义特征驱动的幽默识别模型(CMSOR)

2.1 问题描述

幽默识别任务的具体目标可以描述为:给定幽默识别训练集 $D = \{x_i, y_i\}_{i=1}^N$ 。其中, $x_i = \{w_1, w_2, \dots, w_m\}$ 为输入文本序列, m 为其中包含的单词总数。 $y_i \in [0, 1]$ 为对应的幽默标签, N 为训练集中样本总数。幽默识别任务的目标则是学习到一个映射函数 $f: D \rightarrow y \in [0, 1]$, 以此预测输入文本序列是否为幽默。

2.2 CMSOR 模型框架

本文所提出的基于动态常识推理与多维语义特征的幽默识别(CMSOR)模型如图 1 所示。该模型主要由三个部分组成:情感特征提取层、语义特征提取层、模糊性特征提取层。其中,情感特征提取层主要是考虑到幽默表达中存在大量隐式情感表达,利用外部常识推断文本中的隐式情感表达,充分挖掘文本中的情感特征;语义特征提取层主要是通过计算句子内部词对之间的语义关联来学习文本内部的不一致性特征;模糊性特征提取层主要是利用外部词典捕捉文本中存在歧义性的词汇,通过循环神经网络学习其模糊性特征。最后,将幽默的三个维度特征进行拼接,通过分类器获得文本的幽默预测结果。

2.3 外部常识驱动的情感特征提取

幽默表达与情感有着极大的关联。一些带有强烈感情色彩的词会增加受众对于作者表述的认同感,使得读者的情绪被更为充分地调动,从而达到幽默的效果^[26]。然而幽默内存在的隐式情感表达使得通过外部词典捕捉文本情感特征变得十分困难。为了解决这一问题,本文采用预训练常识推理模块 COMET 根据上下文信息动态推断文本内所蕴含的情感特征。COMET 作为一种常识推理工具,在给定上下文的情况下,能够根据不同的事件关系来推理相应的结果。COMET 是以 Transformer 为基础架构,并在 ATOMIC^[27]数据集上训练得到。该数据集共提供 23 种事件关系,而本文主要采用 [xReact] 这一关系。它的功能是根据上下文推断句子中主语的内心情绪,并以文本形式输出。

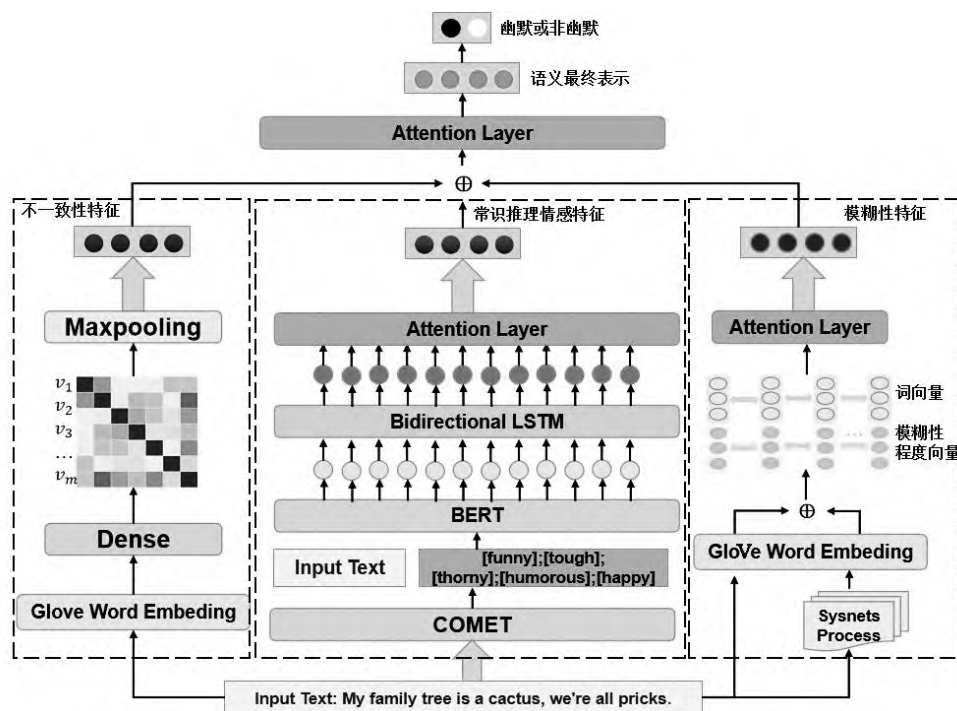


图1 CMSOR 模型框架图

具体地,以幽默文本序列 $x = (w_1, w_2, \dots, w_m)$ 作为输入,COMET 能够根据 x 推理出说话者可能的内心情绪。在这里,本文选择概率最高的前 l 个可能结果,并得到说话者情绪候选集 $K = \{k_1, k_2, \dots, k_l\}$ 。其中, k_i 表示第 i 个情绪词。然后,将初始文本序列 x 与情绪候选集拼接,得到显式情绪增强的幽默文本序列,如式(1)所示。

$$x_e = \{w_1, w_2, \dots, w_m, [\text{SEP}], k_1, k_2, \dots, k_l\} \quad (1)$$

其中, $[\text{SEP}]$ 为句子分割符。本文采用 BERT 对 x_e 进行上下文编码。具体计算如式(2)所示。

$$v_e = \text{BERT}(x_e; W_0) \quad (2)$$

其中, v_e 为编码后得到的句子表示, W_0 为 BERT 的可学习参数。

一方面, BERT 能够有效地捕捉上下文信息,将幽默文本 x 中的单词 w_i 与情绪候选集 K 中的情绪单词 k_i 从语义层面上关联起来,进而有效捕捉文本内的情绪特征。另一方面, BERT 内的多头注意力机制能够为文本中的每个单词赋予不同的权重,通过降低与幽默文本无关的情绪词的权重,来避免引入过多噪声信息。在得到上下文编码后,采用双向长短时记忆神经网络 (Bi-LSTM) 对于上下文语义信息进行进一步学习,最后通过注意力机制获取潜在情感特征 z_e , 其计算公式如式(3)、式(4)所示。

$$u_e = \text{Bi-LSTM}(v_e; W_1) \quad (3)$$

$$z_e = \text{Attention}(u_e; W_2) \quad (4)$$

其中, $u_e \in R^{1 \times p}$ 为输出的幽默文本表示, p 为 Bi-LSTM 的隐藏层维度。 W_1 为 Bi-LSTM 的可学习参数, W_2 为注意力机制的可学习参数。

2.4 基于语义距离的不一致性特征提取

一些语言学研究^[28]认为幽默的本质在于表现出两种不一致的思想或概念。同样的, Raskin 等人^[29]也指出幽默的产生往往借助于一些有意义但含义不同或相反的词语或短语的组合,通过制造错觉或矛盾感而达到幽默的效果。例如,

例1 I am deeply aware that I am a superficial person.

例1中“deeply”可以翻译成“深刻”,“superficial”可以翻译成“肤浅”。这个句子的中文翻译是“我深刻地意识到我是个肤浅的人”,其中“深刻”和“肤浅”有相反的含义,从而达到幽默效果。上述例子也可以说明幽默中的不一致特征具有隐晦和抽象的特点,并与深层次语义关联紧密。从听者角度,需要具有背景知识才能够推断出词汇或者短语之间的隐含关系。因此,需要引入外部知识更好地捕捉幽默的不一致性特征。

具体地,给定一个输入文本序列 $x = (w_1, w_2, \dots, w_m)$, 本文首先通过预训练语言模型将文本序列中的每个词进行向量化表示并得到 $V = (v_1, v_2, \dots, v_m) \in R^{m \times d}$ 。其中 d 表示词向量维度。然后,

针对于 x 中的每个词 w_i , 利用 WordNet 获取其词义特征, 并得到 $H = (h_1, h_2, \dots, h_m) \in \mathbf{R}^{m \times d'}$, d' 表示其词义特征维度。将词义信息 H 与深层次语义信息 V 进行拼接, 得到 $V' = (v'_1, v'_2, \dots, v'_m) \in \mathbf{R}^{m \times (d+d')}$ 。为了计算词级语义不一致性, 首先采用两个平行语义编码器对文本表示 V' 进行压缩。编码器由全连接神经网络实现, 具体计算如式(5)、式(6)所示。

$$\hat{V} = \sigma(W_3 V' + b_2) \quad (5)$$

$$\bar{V} = \sigma(W_4 V' + b_3) \quad (6)$$

其中, \hat{V} 与 \bar{V} 分别表示压缩后的语义信息, W_3 与 W_4 表示两个编码器中的可学习参数, σ 表示激活函数。然后, 对 \hat{V} 与 \bar{V} 进行点积运算, 获得不一致性矩阵 $S = \hat{V} \cdot \bar{V}^T$, 用以刻画幽默文本的内部不一致性特征。

2.5 基于同义词的模糊性特征提取

Reyes 和 Rosso^[30] 认为幽默是一个单词的多个含义令句子产生不同的理解, 借助语义和语境的歧义来产生的。Miller 和 Gurevych^[31] 指出, 模糊性是幽默的关键因素, 是幽默中常见的语言现象。随之 Reyes 等人^[32] 得出结论: 幽默的表达往往伴随着语义的模棱两可。如下例:

例 2 Why did the tomato turn red? Because it saw the salad dressing!

例 3 My trip to the grand canyon cost a hole lot of money and gorged my bank account butte it was worth it.

例 2 中“salad”一词既可以被解释为用于沙拉的一种酱汁, 也可以表示“穿衣服”的意思, 从而导致句子产生两种截然不同的意义来产生幽默效果。例 3 中, 首先“hole”的字面含义为“洞”, 但在口语中也可表示为“大量”或“很多”, 其次“butte”的字面含义为“丘陵”, 但在句中被用作双关词, 与“but”相呼应。句子通过“hole”和“butte”的双关含义, 使得例 3 既可描述为旅行花费了大量的钱, 也可暗示这个花销像一个巨大的洞一样, 吞噬了大量的资金。结合上述例子, 幽默通过词汇的多个含义来创造幽默, 达到幽默效果。由此可见, 模糊性是判断文字是否幽默的重要因素之一, 是幽默文本的重要组成部分。综上所述, 本文为提高幽默识别的性能, 利用外部资源 WordNet 捕获句子中的歧义词。

在 WordNet 数据库中, 名词、动词、形容词和副词都被存储为同义词集合的形式, 每一个同义词集合被称为一个 Synset, 包含一组具有相似意义的单词。不同的 Synset 之间可以通过语义关系和词性关系等边相连接, 这些关系可以帮助人们理解这些单词之间的联系和含义。

针对于输入文本序列 $x = (w_1, w_2, \dots, w_m)$, 首先利用 WordNet 中的同义词集合 Synset 计算每个单词 w_i 的同义集数量 n 。本文认为单词的同义词数量越多, 会导致句子理解存在很多歧义, 从而模糊性程度就会增加, 因此本文将同义集数量最多的词汇设置为最容易出现歧义的词汇, 停用词在句子中不承载实际的语义信息, 因此可以被移除或忽略, 从同义词集合和同义词数量中删除停用词汇及其个数。针对于同义词集的数量, 定义如式(7)所示的规则来描述每个词的模糊程度。

$$c = \begin{cases} 0 & n = -1 \\ 1 & 0 < n \leq 5 \\ 2 & 5 < n \leq 15 \\ 3 & 15 < n \leq 30 \\ 4 & n > 30 \end{cases} \quad (7)$$

由式(7)得到 x 的模糊程度序列 $c = (c_1, c_2, \dots, c_m)$, 其中, 0 表示模糊程度最低, 4 表示模糊程度最高, 对于文本中的停用词, 其模糊程度统一设定为 0。然后, 将该序列 c 进行 one-hot 表示, 得到模糊程度矩阵 $V_c = [c_1, c_2, \dots, c_m] \in \mathbf{R}^{m \times d}$ 。将 V_c 与文本表示 $V = [v_1, v_2, \dots, v_m] \in \mathbf{R}^{m \times d}$ 通过拼接方式进行融合, 并利用模糊特征编码器 G_{fuz} 学习包含模糊特征的文本表示, 该编码器由 Bi-LSTM 及注意力机制实现。其计算如式(8)所示。

$$z_f = G_{fuz}([V_c \oplus V]; W_5) \quad (8)$$

其中, z_f 为模糊性特征表示, W_5 为可学习参数, \oplus 表示拼接操作。

2.6 幽默标签预测以及损失函数

在获得幽默文本的情感特征 z_e 、不一致性特征 $z_s = \text{MaxPooling}(S)$ 以及模糊性特征 z_f 之后, 将三种特征通过拼接方式进行融合, 得到多维度融合幽默特征 $z = z_e \oplus z_s \oplus z_f$ 。通过注意力机制进一步学习三种特征之间的内在关联, 具体计算如式(9)所示。

$$Z = \text{Attention}(z; W_6) \quad (9)$$

其中, W_6 表示注意力机制层的可学习参数。在此基础上, 将其输入到由全连接层构成的幽默分类器 f_h 中, 获得文本 x 的幽默标签预测。具体计算如

式(10)所示。

$$\hat{y} = f_h(\mathbf{Z}; \mathbf{W}_7) \quad (10)$$

其中, \mathbf{W}_7 表示可学习的参数矩阵, \hat{y} 表示幽默预测结果。

最后, CMSOR 在分类中采用交叉熵 (Cross Entropy) 作为损失函数。其损失计算如式 (11) 所示。

$$L = -\frac{1}{N} \left(\sum_i (\hat{y}_i \log y_i + (1 - \hat{y}_i) \log (1 - y_i)) \right) \quad (11)$$

3 实验与分析

本节首先详细描述数据集、实验数据和实验设置, 然后对比了基线方法和本文提出的 CMSOR 方法的幽默识别性能, 最后结合对比实验以及消融实验结果来讨论模型的性能, 并验证本文模型方法的有效性。

3.1 数据集

为了证明方法的有效性, 本文实验中使用了三个公开的数据集, 其统计信息如表 2 所示。具体介绍如下:

Pun of The Day^[33]: 该数据集的构建是 Yang 等人通过在互联网上收集幽默文本而完成的, 包括了各种类型的幽默, 如双关语、笑话、俏皮话等等。为确保数据的准确性和可靠性, 通过人工标注和质量控制的方式对数据进行了筛选和整理。该数据集目前广泛使用于幽默识别中。

SemEval 2021 Task7-1a^[34]: 该任务是一项国际评测, Task7 子任务一是识别文本是否为幽默文本, 该数据集可以用来进行幽默检测, 本文利用 Task7 子任务一涉及数据来判断是否为幽默文本。

ColBERT^[22]: 该数据集是一个大规模的幽默数据集, 它包含了 20 万个来自网络的英文幽默文本, 其中 10 万正样本由 Reddite 收集得到, 另外 10 万负样本来源于新闻头条。

表 2 数据集统计信息

数据集	正样本	负样本
Pun of The Day	2 403	2 403
SemEval 2021 Task7-1a	5 547	3 457
ColBERT	100 000	100 000

3.2 实验数据与设置

实验在 Python 3.7 和 Keras 2.2.4 环境下进行。对于本文提出的 CMSOR 模型, 其中常识知识层本文采用 12 层的 BERT-base-cased^① 作为预训练语言模型编码, 其中向量维度为 768, 共 110M 个参数; 语义特征提取以及模糊性特征提取采用 GloVe, 维度为 100, 词嵌入在训练的过程中固定, 不在词汇表中出现的单词使用 $(-0.01, 0.01)$ 上的平均分布随机初始化; 使用 WordNet 获取单词同义词集合; Bi-LSTM 的神经元数量为 128; Dropout 为 0.3; Batch 大小为 64; 模型采用 Adam Optimization 优化算法更新模型参数; 采用了学习率衰减和早停机制以防止过拟合现象。此外, 采用准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F_1 值 (F_1 -Score) 作为实验结果的评价指标, 并且所有实验均进行五倍交叉验证, 取平均值作为实验结果。

3.3 基线方法

本文采用如下基线模型进行对比:

(1) LSTM^[35]: 通过经典 LSTM 模型提取幽默特征进行幽默识别。

(2) Bi-LSTM: 利用可以更好地捕捉双向语义依赖关系的 Bi-LSTM 模型。

(3) Bi-LSTM+ATT: 使用 Bi-LSTM 模型结合注意力机制提取幽默特征进行幽默识别。

(4) CNN: 采用卷积神经网络 (Convolutional Neural Network, CNN) 获取幽默语句的潜在语义及模糊性特征进行幽默识别。

(5) BERT^[36]: 使用预训练 BERT 模型在幽默数据集上进行微调。

(6) IEANN^[21]: 通过结合内部及外部注意力神经网络构建两种注意力机制, 以捕捉幽默文本中的不一致性和模糊性特征。

(7) ABML^[24]: 通过联合幽默和双关语检测的多任务学习模型进行幽默识别。

(8) ANPLS^[25]: 通过结合发音、词汇和句法幽默特征的注意力网络, 提取幽默特征进行幽默识别。

3.4 对比实验

实验结果如表 3 所示, 对实验结果进行具体分析可以得到如下结论:

① <https://huggingface.co/bert-base-cased>

表 3 对比实验结果

Model	Dataset											
	Pun of the Day				SemEval 2021 Task 7-1a				ColBERT			
	ACC/%	P/%	R/%	F ₁ /%	ACC/%	P/%	R/%	F ₁ /%	ACC/%	P/%	R/%	F ₁ /%
LSTM	84.97	84.02	84.57	84.29	83.30	83.06	81.34	82.19	93.60	93.82	94.05	93.93
Bi-LSTM	86.11	85.13	85.87	85.50	84.90	87.79	87.64	87.71	94.07	94.80	93.19	94.08
Bi-LSTM+ATT	86.94	87.95	84.13	86.00	84.70	87.62	87.48	87.55	95.48	96.01	94.84	95.42
CNN	86.42	83.18	91.56	87.17	86.15	87.20	90.32	88.73	94.40	93.18	95.81	94.45
BERT	90.50	88.75	91.80	90.25	91.78	91.66	92.62	92.14	95.55	95.57	95.47	95.52
IEANN	92.24	91.14	92.25	91.69	91.03	91.32	92.10	91.71	94.92	95.33	93.87	94.59
ABML	93.18	92.45	92.07	92.26	92.20	91.92	92.77	92.34	94.42	94.39	94.16	94.27
ANPLS	92.94	93.00	92.55	92.79	92.06	92.38	93.07	92.72	94.39	94.82	95.07	94.94
CMSOR	94.56	93.47	92.61	93.04	92.15	92.67	93.40	93.03	96.23	95.98	96.40	96.19

(1) 本文提出的 CMSOR 方法在三个数据集上均取得了最好的结果,在三个数据集上的 F_1 值相比于现存的最优结果分别提升了 0.25%、0.31%、0.67%,证明了从情感、不一致性以及模糊性三个维度构建幽默语义并应用于幽默识别是有效的。

(2) 从表中可以看出,相比于基于 CNN 的幽默识别方法,基于 Transformer 的方法(BERT 以及 CMSOR)在四项评价指标上有明显性能提升,这说明 Transformer 能够通过全局注意力机制更好地捕捉幽默文本的上下文信息。

(3) CMSOR 方法能够通过深度神经网络结构,在外部知识驱动下,自动构建幽默特征,相比于 BERT 在 Pun of The Day 数据集上取得了明显的性能提升(F_1 值提高 2.79%)。这也验证了深度学习模型结合外部知识能够在幽默理论约束下学习到幽默相关特征。

(4) 相比于基于 RNN 的方法,基于 CNN 的方法在 Pun of The Day 和 SemEval 数据集上 F_1 值取得了明显的提升,如在 SemEval 数据集上, CNN 相比于 Bi-LSTM+ATT 在 F_1 值提高了 1.18%。这说明幽默表达可能与局部语义信息(Ngram)有着一定的关联。

(5) 与采用情感词典捕捉文本内部情感信息的 IEANN 相比,本文提出的 CMSOR 方法在 Pun of The Day、SemEval 以及 ColBERT 数据集上 F_1 值分别提升了 1.35%、1.32%和 1.6%。这说明利用动态外部常识信息能够更准确地推断文本内部情感。

(6) ABML 模型在 Pun of The Day 和 SemEval

数据集上 ACC 值高于 IEANN 和 ANPLS。ACC 值达到最高。ABML 模型不仅考虑双关语的特点,还考虑了幽默和双关语之间共同的潜在语义信息。这意味着模型能够更好地理解双关语的双重含义,并将其与幽默特征联系起来,有效地增强模型对幽默的识别能力。

3.5 消融实验

为了验证 CMSOR 中不同组件的有效性,本文在三个数据集上进行消融实验,并设计以下模型变体: CMSOR-C 表示仅使用情感特征;CMSOR-I 表示仅使用语义不一致性特征;CMSOR-A 表示仅使用模糊性特征;CMSOR-CI 表示融合情感特征和语义不一致性特征;CMSOR-CA 表示融合情感特征和模糊性特征;CMSOR-IA 表示融合语义不一致性特征和模糊性特征。

三个数据集上的消融实验结果如表 4 所示。从表中可以得到如下结论:

(1) 当分别移除情感特征(CMSOR-IA)、模糊特征(CMSOR-CI)以及不一致性特征(CMSOR-CA)之后,模型在 SemEval 2021 Task7-1a 数据集上的四项指标均有明显下降(F_1 值分别下降 3.26%、1.82%、1.45%),这说明三种情感特征在幽默识别任务中的有效性。然而,在 Pun of The Day 数据集上,当移除情感特征后,模型在召回率 R 上有提升,这可能是因为 BERT 在学习情感增强的文本表示时,将错误的情绪信息融入到语义表示当中,所以导致该指标下降。而在 ColBERT 数据集上,精

表 4 消融实验结果

Model	Dataset											
	Pun of the Day				SemEval 2021 Task 7-1a				ColBERT			
	ACC/%	P/%	R/%	F ₁ /%	ACC/%	P/%	R/%	F ₁ /%	ACC/%	P/%	R/%	F ₁ /%
BiLSTM	86.11	85.13	85.87	85.50	84.90	87.79	87.64	87.71	93.96	93.82	94.05	93.93
CMSOR-C	86.53	85.71	86.09	85.90	86.60	88.24	90.11	89.17	94.07	94.80	93.19	94.28
CMSOR-I	86.32	86.12	85.00	85.56	85.70	86.99	90.24	88.59	94.45	96.66	92.00	94.08
CMSOR-A	91.48	96.27	86.20	90.96	86.30	87.94	90.08	89.00	95.65	96.30	94.90	95.59
CMSOR-CI	91.81	92.24	90.43	91.33	91.61	91.11	91.30	91.21	95.72	94.90	96.59	95.74
CMSOR-CA	92.23	91.22	92.61	91.91	91.92	90.99	92.17	91.58	95.86	94.69	97.13	95.89
CMSOR-IA	92.95	92.06	93.26	92.66	87.45	89.59	89.95	89.77	96.21	96.77	95.61	96.19
CMSOR	94.56	93.47	92.61	93.04	92.15	92.67	93.40	93.03	96.23	95.98	96.40	96.19

确率 P 得到了提升,这可能是因为情感特征与其他特征存在冗余,即它们所携带的信息在一定程度上是重叠的。当移除情感特征时,模型可能更加依赖于其他更为关键的特征,从而提高了精确率。

(2) 当只保留模糊性特征的时候,模型在 Pun of The Day 和 ColBERT 数据集上的表现相比于 CMSOR 下降得最少,在 Pun of The Day 数据集上精确率 P 得到了提升,这说明模糊性特征在构建幽默语义的过程中相比于情感特征以及不一致性特征更加重要。然而,对于 SemEval 2021 Task7-1a 数据集,情感特征更加重要。

4 参数分析

图 2 展示了不用数量的常识信息对于模型性能

的影响。从图中可以观察到,在 Pun of The Day 和 ColBERT 数据集上,当知识数量为 1 时,模型效果最差。随着候选知识数量的不断增加,模型的表现逐渐提升,并且在 $l = 5$ 时取得最好的结果。这说明有效处理隐式情感表达对于 CMSOR 建模幽默语义具有重要作用,并且显式情感信息的增加会提升模型对于文本情感特征的捕捉效果。对于 SemEval 2021 Task7-1a 数据集而言,变化趋势与其他两个数据集不同。随着知识数量的增加,模型表现在性能略微提升之后,呈现出下降趋势,并且在 $l = 4$ 时取得最差的结果,但是 $l = 5$ 时结果最优。这可能是因为当将知识数量增加到 5 时,一些样本的隐式情感表达才能够被 COMET 有效推理出来。

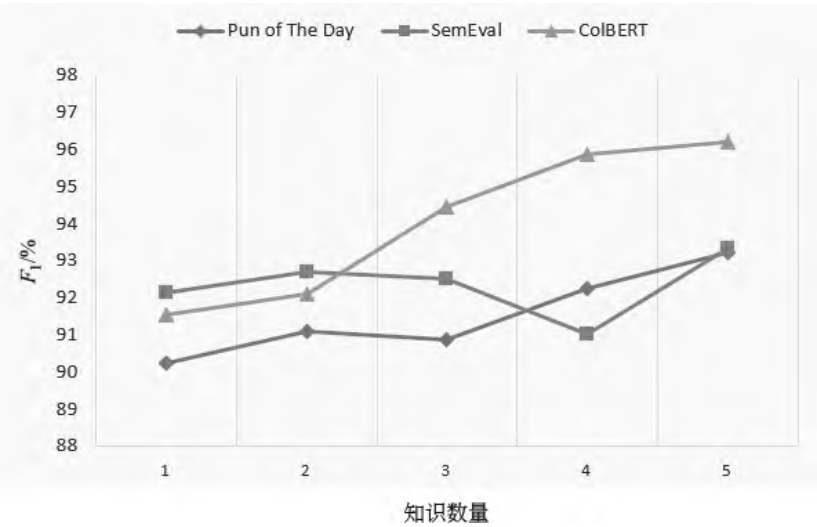


图 2 不同数量的知识候选对模型性能的影响

5 总结

针对于现有幽默识别方法没有充分捕捉文本内部的情感特征,忽略了幽默文本中的隐式情感表达这一问题,本文提出一种动态常识与多维语义特征驱动的幽默识别方法 CMSOR。该方法首先利用外部常识信息从文本中动态推理出说话者的隐式情感表达,然后引入外部词典 WordNet 计算文本内部词级语义距离,进而捕捉不一致性,同时计算文本的模糊性特征。最后,根据上述三个特征维度构建幽默语义,实现幽默识别。本文在三个公开数据集上进行实验,结果表明本文所提方法 CMSOR 相比于当前基准模型性能有明显提升。未来,本文将尝试把常识信息应用到幽默生成、多模态幽默识别等任务当中。

参考文献

- [1] RITCHIE G. Can computers create humor? [J]. AI Magazine, 2009, 30(3): 71-79.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [3] SULLS J M. A two-stage model for the appreciation of jokes and cartoons [J]. An Information-processing Analysis. The Psychology of Humor: Theoretical Perspectives and Empirical Issues, 1972, 1: 81-100.
- [4] CHAUHAN D S, et al. A sentiment and emotion aware multimodal multiparty humor recognition in multilingual conversational setting [C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 6752-6761.
- [5] LIU LZH, DONGHAI ZH, et al. Modeling sentiment association in discourse for humor recognition [C]//Proceedings of 56th Annual Meeting of the Association for Computational Linguistics, 2018: 586-591.
- [6] LI D, et al. HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media [J]. Information Processing & Management, 2020, 57(6): 102290.
- [7] BOSSELUUT A, RASHKIN H, SAP M, et al. COMET: Commonsense transformers for automatic knowledge graph construction [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 4762-4779.
- [8] MILLER G A. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.
- [9] MIHALCEA R, STRAPPARVA C. Making computers laugh: investigations in automatic humor recognition [C]//Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, 2005: 531-538.
- [10] MIHALCEA R, STRAPPARVA C, PULMAN S. Computational models for incongruity detection in humour [C]//Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics, 2010: 364-374.
- [11] YANG D, LAVIE A, DYER C, et al. Humor recognition and humor anchor extraction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015: 2367-2376.
- [12] MORALES A, ZHAI C. Identifying humor in reviews using background text sources [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017: 492-501.
- [13] CATTLE A, MA X. Recognizing humour using word associations and humour anchor extraction [J]. Association for Computational Linguistics, 2018: 1849-1858.
- [14] BERTERO D, FUNG P. A long short-term memory framework for predicting humor in dialogues [C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 130-135.
- [15] CUZA C M, BUENO R O, ROSSO P, et al. UOUPV: Deep linguistic humor detection in spanish social media [C]//Proceedings of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages Colocated with 34th Conference of the Spanish Society for Natural Language Processing, 2018: 204-213.
- [16] BLINOV V, BOLOTOVA V, BRASLAVSKI P. Large dataset and language model fine-tuning for humor recognition [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 4027-4032.
- [17] KAO J, LEVY R, GOODMAN N D. A computational model of linguistic humor in puns [J]. Cognitive Science, 2016: 1270-1285.
- [18] WELLER O, SEPPI K. Humor detection: A transformer gets the last laugh [C]//Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 3612-3616.
- [19] HASAN M K, RAHMAN W, ZADEH A, et al. UR-FUNNY: A multimodal language dataset for understanding humor [J]. arXiv Preprint arXiv: 1904.06618, 2019.

- [20] DIAO Y, YANG L, ZHANG D, et al. Homographic puns recognition based on latent semantic structures [C]//Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing, 2018;565-576.
- [21] FAN X, LIN H, YANG L, et al. Humor detection via an internal and external neural network[J].Neuro-computing, 2020,394(1):105-111.
- [22] ANNAMORADNEJAD I, ZOGHI G. ColBERT: Using BERT sentence embedding for humor detection [J].2004.12765v2, 2021.
- [23] ZHANG T Y, ZHANG SH W, XU B, et al.结合标签转移关系的多任务笑点识别方法[C]//Proceedings of the 20th Chinese National Conference on Computational Linguistics. Huhhot, China: Chinese Information Processing Society of China, 2021;238-247.
- [24] REN L, XU B, LIN H, et al. ABML: Attention-based multi-task learning for jointly humor recognition and pun detection[J].Soft Computing, 2021, 25(22):14109-14118.
- [25] REN L, XU B, LIN H, et al. An attention network via pronunciation, lexicon and syntax for humor recognition[J].Applied Intelligence, 2022, 52(3): 2690-2702.
- [26] 樊小超, 杨亮, 林鸿飞, 等. 基于多维潜在语义特征的幽默识别[J].中文信息学报, 2021, 35(8): 38-46.
- [27] SAP M, et al. Atomic: An atlas of machine commonsense for if-then reasoning [C]//Proceedings of the AAAI conference on Artificial Intelligence, 2019, 33(1):3027-3035.
- [28] LEFCOURT H M. Humor: The psychology of living buoyantly.[J].Contemporary Sociology, 2001,31(1): 35-48.
- [29] RASKIN V. Semantic mechanisms of humor[M]//D. Reidel Pub. Co., 1984;325-335.
- [30] REYES A, ROSSO P. Making objective decisions from subjective data: Detecting irony in customer reviews[J].Decision Support Systems, 2012, 53(4): 754-760.
- [31] MILLER T, GUREVYCH I. Automatic disambiguation of English puns [C]//Proceedings of the International Joint Conference on Natural Language Processing. Association for Computational Linguistics, 2015;719-729.
- [32] REYES A, ROSSOP, BUSCALDI D. From humor recognition to irony detection: The figurative language of social media[J].Data & Knowledge Engineering, 2012, 74:1-12.
- [33] YANG D, LAVIE A, DYER C, et al. Humor recognition and humor anchor extraction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015;2367-2376.
- [34] MEANEY J A, et al. Semeval 2021 task 7: Hahackathon, detecting and rating humor and offense [C]//Proceedings of the 15th International Workshop on Semantic Evaluation. 2021;105-119.
- [35] GRAVES A, ALEX G. Long short-term memory [J]. Supervised Sequence Labelling with Recurrent Neural Networks, 2012: 37-45.
- [36] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. Association for Computational Linguistics, 2019, 6:4171-4186.



吐妮可·吐尔逊(1997—), 硕士研究生, 主要研究领域为自然语言处理、情感分析。
E-mail: Tunuh@mail.dlut.edu.cn



林鸿飞(1962—), 博士, 教授, 主要研究领域为自然语言处理、信息挖掘。
E-mail: hflin@dlut.edu.cn



闵昶荣(1994—), 通信作者, 博士, 主要研究领域为自然语言处理、情感分析。
E-mail: mcer19940816@gmail.com