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[a], 林鸿飞*[a], 张冬瑜[b], 杨亮[a], 闵昶荣[a]
[a]大连理工大学，计算机科学与技术学院，大连，116024
[b]大连理工大学，软件学院，大连，116620
(Tunuh,11909060)@mail.dlut.edu.cn,(hflin,zhangdongyu,liang)@dlut.edu.cn

摘要

随着社交媒体的飞速发展，幽默识别任务在近年来受到研究者的广泛关注。该任务的目标是判断给定的文本是否表达幽默。现有的幽默识别方法主要是在幽默产生理论的支撑下，利用规则或者设计神经网络模型来提取多种幽默相关特征，比如不一致性特征、情感特征以及语音特征等等。这些方法一方面说明情感信息在建模幽默语义当中的重要地位，另一方面说明幽默语义的构建依赖多个维度的特征。然而，这些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，影响幽默识别的准确性。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**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有明显提升。

关键词： 幽默识别 ； 常识推理 ； 模糊理论 ； 注意力机制

Humor Recognition based on Dynamically Commonsense Reasoning and Multi-Dimensional Semantic Features

Tuerxun·Tunike[a],Hongfei Lin* [a], Dongyu Zhang[b],Liang Yang[a], Changrong Min[a]
[a]School of Computer Science and Technology,Dalian University of Technology,Dalian,116024
[b]School of Software Technology,Dalian University of Technology,Dalian,116620
(Tunuh,11909060)@mail.dlut.edu.cn,(hflin,zhangdongyu,liang)@dlut.edu.cn

Abstract

With the rapid development of social media,humor recognition has been a popular topic in the community of NLP.The goal of this task is to discriminate whether a given text expresses humor. Existing humor recognition methods mainly rely on the support of humor-centered theory, and use rules or designed neural network architectures to extract various humor-specific features, such as inconsistency features, emotional features, and linguistic features, etc. These methods indicate the importance of emotional information for modeling humor semantics, and also show that the construction of humor semantics depends on multi-dimensional features. However, these methods do not fully capture such emotional features within the text, ignoring implicit emotional expressions in humorous texts, which affects the accuracy of humor recognition. Therefore, we propose a novel approach named **CMSOR**, which is based on dynamic commonsense and

* 通讯作者

第二十二届中国计算语言学大会论文集，第328页-第340页，哈尔滨，中国，2023年8月3日至5日。

multi-dimensional semantic features for humor recognition. Specifically, it first makes use of external commonsense to infer latent emotions of speakers from the given text, and then leverage WordNet lexicon to calculate semantic distances from the word level, aiming to capture inconsistent features. This lexicon is also used to calculate the ambiguous features of the text. Eventually, we make use of such three kinds of humor-specific features to construct humor semantics. We conduct experiments over three publicly available benchmarks. The experimental results demonstrate that the proposed **CMSOR** is superior to the state-of-the-art baselines.

Keywords: Humor detection , Commonsense reasoning , Ambiguity theory , Attention mechanism

1 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其他任务，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(Ritchie, 2009)、宽慰论(LeCun et al., 2015)以及乖讹论(Suls, 1972)。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；乖讹论又称不一致性理论，它的表达方式通常会包含一些出人意料的不一致性，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。比如，Chauhan等人(Chauhan et al., 2022)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)的多任务框架来检测幽默。Liu等人(Liu et al., 2018)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的解决对话幽默识别问题。Li等人(Li et al., 2020)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Bi-LSTM模型结合注意力网络的方法，更好地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“get fired”来表达。这种方式称之为隐式情感表达。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，这降低了模型识别文本幽默的能力。

从认知角度，理解这些隐式情绪表达需要不仅需要结合上下文信息，还有充分利用外部常识。尽管现有的预训练语言模型（PLM）能够高效的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(**C**ommonsense and **M**ulti-dimensional **S**emantics based Humor Detect**or**)，简写为**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 (Bosselut et al., 2019)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

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

Table 1: 幽默样本以及包含的情绪信息

的文本表示。同时，利用外部词典WordNet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。本文做出的贡献总结如下：

（1）本文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）本文在Pun of the Day、SemEval21以及ColBERT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，说明了该方法的有效性。

本文组织结构如下：第2章主要介绍幽默识别相关工作。第3章介绍文本所提出的**CMSOR**模型。第4章主要介绍实验设置以及实验结果分析。第5章为总结与展望。

2 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的任务。早期的幽默识别方法主要是基于特征工程，利用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。比如，Mihalcea和Strapparova(Mihalcea and Strapparava, 2005)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(Mihalcea et al., 2010)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(Yang et al., 2015)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(Morales and Zhai, 2017)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(Cattle and Ma, 2018)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，无法对于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。

随着计算能力的进步以及社交媒体数据的爆炸性增长，深度学习在不同领域被广泛用以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。比如，Bertero等人(Bertero and Fung, 2016)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期记忆网络（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用于识别笑点。Buenod等人(Ortega-Bueno et al., 2018)针对西班牙推文结合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(Blinov et al., 2019)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。Justine T. Kao等人(Kao et al., 2016)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(Weller and Seppi, 2019)使用Transformer架构识别幽默。Hasan等人(Hasan et al., 2019)使用循环神经网络针进行多模态幽默识别。Diao等人提出(Diao et al., 2018)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(Fan et al., 2020) 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(Annamoradnejad and Zoghi, 2020)改进Bert模型在自创建的幽默数据集ColBERT上进行实验，证实了提出模型能够有效的检测幽默。Zhang等人(Zhang et al., 2021)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(Ren et al., 2021)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(Ren et al., 2022)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，本文为了解决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。

3 模型

3.1 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$\mathcal{D} = \{x_i, y_i\}_{i=1}^{N}$。其中$x_i = (w_1, w_2, \ldots, w_m)$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$f : \mathcal{D} \rightarrow y \in [0, 1]$，以此预测输入文本序列是否为幽默。

3.2 模型整体框架

本文所提出的幽默识别方法**CMSOR**结构如图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。

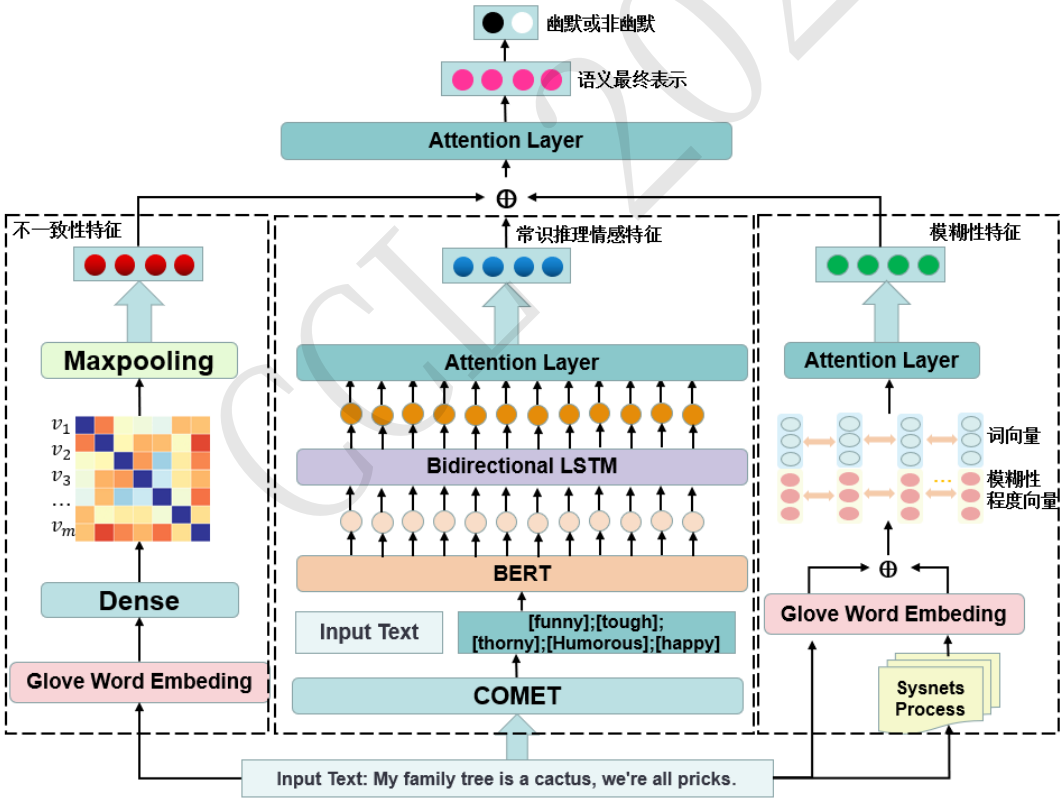

Figure 1: **CMSOR**模型结构图

3.3 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的调动，从而达到幽默的效果(樊小超et al., 2021)。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(Bosselut et al., 2019)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$^{20}_{20}$(Sap et al., 2019)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。

具体地，以幽默文本序列$x = (w_1, w_2, \ldots, w_m)$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$K = \{k_1, k_2, \ldots, k_l\}$。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：

$$x_e = \{w_1, w_2, \ldots, w_m, [\mathbf{SEP}], k_1, k_2, \ldots, k_l\} \tag{1}$$

其中，[**SEP**]为句子分割符。然后，本文采用BERT对于x_e进行上下文编码。其计算公式如下：

$$v_e = \mathbf{BERT}\left(x_e; \mathbf{W_0}\right) \tag{2}$$

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

一方面，BERT能够有效的捕捉上下文信息，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，其计算公式如下：

$$u_e = \mathbf{Bi-LSTM}\left(v_e; \mathbf{W_1}\right) \tag{3}$$

$$z_e = \mathbf{Attention}\left(u_e; \mathbf{W_2}\right) \tag{4}$$

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

3.4 基于语义距离的不一致性特征提取

一些语言学研究(Lefcourt, 2001; Paulos, 2008)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(Raskin, 1979)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。比如：

例*3.1*：*I am deeply aware that I am a superficial person.*

例3.1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“我深刻的意识到我是个肤浅的人”，其中“深刻”和“肤浅”有相反的含义，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$x = (w_1, w_2, \ldots, w_m)$，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$\mathbf{V} = [v_1, v_2, \ldots, v_m] \in \mathbb{R}^{m\times d}$。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(Miller, 1995)获取其词义特征，并得到$\mathbf{H} = [h_1, h_2, \ldots, h_m] \in \mathbb{R}^{m\times d'}$，$d'$表示其词义特征维度。将词义信息$\mathbf{H}$与深层次语义信息$\mathbf{V}$进行拼接，得到$\mathbf{V'} = [v'_1, v'_2, \ldots, v'_m] \in \mathbb{R}^{m\times(d+d')}$。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于文本表示V'进行压缩。编码器由全连接神经网络实现。具体计算如下：

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

$$\overline{\mathbf{V}} = \sigma(\mathbf{W_4}\mathbf{V'} + \mathbf{b_3}) \tag{6}$$

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

3.5 基于同义词的模糊性特征提取

Reyes和Rosso(Reyes and Rosso, 2012)认为幽默是一个单词的多个含义对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(Miller and Gurevych, 2015)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(Reyes et al., 2012)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*3.2*: *Why did the tomato turn red? Because it saw the salad dressing!*

例*3.3*: *My trip to the grand canyon cost a hole lot of money and gorged my bank account butte it was worth it.*

例3.2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3.3中，首先"*hole*"字面含义为"洞"，但在口语中也可表示为"大量"或"很多"，其次"*butte*"在字面含义为"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，使例3.3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$x = \{w_1, w_2, \ldots, w_m\}$，首先利用WordNet中的同义词集合Synset计算每个w_i的同义集数量n。本文认为单词的同义词数目越多，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，定义如下规则来描述每个词的模糊程度：

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

得到x的模糊程度序列$C = \{c_1, c_2, \ldots, c_m\}$，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$\mathbf{V_c} = [c_1, c_2, \ldots, c_m] \in \mathbb{R}^{m \times d}$。将$\mathbf{V_c}$与文本表示$\mathbf{V} = [v_1, v_2, \ldots, v_m] \in \mathbb{R}^{m \times d}$通过拼接方式进行融合，并利用模糊特征编码器$\mathcal{G}_{\mathbf{fuz}}$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。其计算公式如下：

$$z_f = \mathcal{G}_{\mathbf{fuz}}\left([\mathbf{V_c} \oplus \mathbf{V}]; \mathbf{W_5}\right) \tag{8}$$

其中，z_f为模糊性特征表示，p'为Bi-LSTM的隐藏层维度。$\mathbf{W_5}$为可学习参数,$\oplus$表示拼接操作。

3.6 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$z_s = \mathbf{MaxPooling}(\mathbf{S})$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，具体计算如下：

$$Z = \mathrm{Attention}(z; \mathbf{W_6}) \tag{9}$$

其中，$\mathbf{W_6}$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。具体计算公式如下：

$$\hat{y} = f_h(Z; \mathbf{W_7}) \tag{10}$$

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。其损失计算如下：

$$\mathcal{L} = -\frac{1}{\mathrm{N}} \sum_{i}^{\mathrm{N}} \left(\hat{y}_i \log y_i + (1 - \hat{y}_i) \log (1 - y_i)\right) \tag{11}$$

4 实验与分析

在本节中，将从以下四个部分介绍实验的细节：数据集、实验数据与设置、对比试验、消融实验。

4.1 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** (Yang et al., 2015)：这个数据集的构建是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 Task 7-1a** (Meaney et al., 2021)：该任务是一项国际评测，Task 7子任务一是识别文本是否为幽默文本，该数据集可以用来幽默检测，本文利用Task 7子任务一涉及数据来判断是否为幽默文本。

- **ColBERT** (Annamoradnejad and Zoghi, 2020)：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。

数据集	正样本	负样本
Pun of The Day	2403	2403
SemEval 2021 Task 7-1a	5547	3453
ColBERT	100,000	100,000

Table 2: 数据集统计信息

4.2 实验数据与设置

实验在python3.7和Kreas2.2.4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 [0] 作为预训练语言模型编码输和知识，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100，词嵌入在训练的过程中固定，不在词汇表中出现的单词词使用(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 Optimzation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。此外采用准确度（Accuracy）、精确率(Precision)、召回率(Recall) 和F1值(F1-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。

[0]https://huggingface.co/bert-base-cased

4.3 对比试验

本文采用如下基线模型进行对比:

- **LSTM** (Graves and Graves, 2012):通过经典LSTM 模型提取幽默特征进行幽默识别。
- **Bi-LSTM**:利用可以更好的捕捉双向语义依赖关系的Bi-LSTM模型。
- **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。
- **CNN**:采用CNN获取幽默语句的潜在语义及模糊性特征进行幽默识别。
- **CNN+F+HN** (Chen and Soo, 2018):采用了融合人工特征的CNN和highway神经网络模型。
- **BERT** (Devlin et al., 2018):使用预训练BERT模型在幽默数据集上进行微调。
- **IEANN** (Fan et al., 2020):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。
- **ABML** (Ren et al., 2021):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。
- **ANPLS** (Ren et al., 2022):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。

Dataset	PUN OF THE DAY			
Model	ACC	P	R	F1
LSTM	84.97%	84.02%	84.57%	84.29%
Bi-LSTM	86.11%	85.13%	85.87%	85.50%
Bi-LSTM+ATT	86.94%	87.95%	84.13%	86.00%
CNN	86.42%	83.18%	91.56%	87.17%
CNN+F+HN*	89.40%	86.60%	94.00%	90.10%
BERT	90.50%	88.75%	91.80%	90.46%
IEANN	92.24%	91.14%	92.25%	91.69%
ABML	93.18%	92.45%	92.07%	92.26%
ANPLS	92.94%	93.00%	92.55%	92.79%
CMSOR	**94.56%**	**93.47%**	**92.61%**	**93.24%**

Table 3: Pun of The Day 数据集上实验结果，*表示结果引用自对应论文，加粗表示最优实验结果。

Dataset	SemEval 2021 Task 7-1a			
Model	Acc	P	R	F1
LSTM	83.30%	83.06%	81.34%	82.44%
Bi-LSTM	84.90%	87.79%	87.64%	87.71%
Bi-LSTM+ATT	84.70%	87.62%	87.48%	87.55%
CNN	86.15%	87.20%	90.32%	89.02%
BERT	91.78%	93.29%	92.62%	92.14%
IEANN	91.03%	91.32%	92.10%	91.71%
ABML	**92.20%**	91.92%	92.77%	92.34%
ANPLS	92.06%	92.38%	93.07%	92.72%
CMSOR	92.15%	**92.67%**	**93.40%**	**93.34%**

Table 4: SemEval 2021 Task 7-1a数据集上实验结果，加粗表示最优实验结果。

Dataset	Colbert			
Model	Acc	P	R	F1
LSTM	93.60%	93.82%	94.05%	93.93%
Bi-LSTM	94.07%	94.80%	93.19%	94.08%
Bi-LSTM+ATT	95.48%	96.01%	94.84%	95.42%
CNN	94.40%	93.18%	95.81%	94.45%
BERT	95.55%	95.57%	95.47%	95.52%
IEANN	94.92%	95.33%	93.87%	94.59%
ABML	94.42%	94.39%	94.16%	94.27%
ANPLS	94.39%	94.82%	95.07%	94.94%
CMSOR	**96.23%**	**95.98%**	**96.40%**	**96.19%**

Table 5: ColBERT数据集上实验结果，加粗表示最优实验结果。

实验结果如表3，4，5所示，从表中可以得到如下结论：(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F1值相比于现存的最优结果分别提升了0.45%、0.62%、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或者RNN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，相比于人工提取幽默特征（CNN+F+HN），取得了明显的提升（F1值提高3.14%）。这也验证了深度学习模型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的方法在三个数据集上的F1值取得了明显的提升，比如在Pun of The Day数据集上，CNN+F+HN相比于BILSTM+ATTEN在F1值上提高了4.46%。这说明幽默表达可能与局部语义信息（N-gram）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，**CMSOR**在F1值上有明显提升（1.6%），这说明利用动态外部常识信息能够更准确的推断文本内部情感。(6)ABML模型在三个数据集上相比较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，有效的增强模型对幽默的识别能力。

4.4 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。

Model	ACC	P	R	F1
BILSTM	86.11%	85.13%	85.87%	85.50%
CMSOR-C	86.53%	85.71%	86.09%	85.90%
CMSOR-I	86.32%	86.12%	85.00%	85.56%
CMSOR-A	91.48%	96.27%	86.20%	90.96%
CMSOR-CI	91.81%	92.24%	90.43%	91.33%
CMSOR-CA	92.23%	91.22%	92.61%	91.91%
CMSOR-IA	92.95%	92.06%	**93.26%**	92.66%
CMSOR	**94.56%**	**93.47%**	92.61%	**93.24%**

Table 6: Pun of The Day数据集消融实验，加粗表示最优实验结果。

三个数据集上的消融实验结果分别如表6，7，8所示。从三个表中可以得到如下结

Model	ACC	P	R	F1
BILSTM	84.90%	87.79%	87.64%	87.71%
CMSOR-C	86.60%	88.24%	90.11%	89.17%
CMSOR-I	85.70%	86.99%	90.24%	88.59%
CMSOR-A	86.30%	87.94%	90.08%	89.00%
CMSOR-CI	91.61%	91.11%	91.30%	91.21%
CMSOR-CA	91.92%	90.99%	92.17%	91.58%
CMSOR-IA	87.45%	89.59%	89.95%	89.77%
CMSOR	**92.15%**	**92.67%**	**93.40%**	**93.34%**

Table 7: SemEval 2021 Task 7-1a数据集消融实验，加粗表示最优实验结果。

Model	ACC	P	R	F1
BILSTM	93.96%	93.82%	94.05%	93.93%
CMSOR-C	94.07%	94.80%	93.19%	94.28%
CMSOR-I	94.45%	96.66%	92.00%	94.08%
CMSOR-A	95.65%	96.30%	94.90%	95.59%
CMSOR-CI	95.72%	94.90%	96.59%	95.74%
CMSOR-CA	95.86%	94.69%	97.13%	95.89%
CMSOR-IA	96.21%	**96.77%**	95.61%	**96.19%**
CMSOR	**96.23%**	95.98%	**96.40%**	**96.19%**

Table 8: ColBERT数据集消融实验，加粗表示最优实验结果。

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 Task 7-1a数据集上的四项指标均有明显下降（F1值分别下降3.57%，2.13%，1.76%），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上有了提升，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。同时，这种情况还出现在ColBERT数据集上，原因同上。(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**下降的最少，这说明模糊性特征在构建幽默语义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 Task 7-1a数据集，情感特征更加重要。

5 参数分析

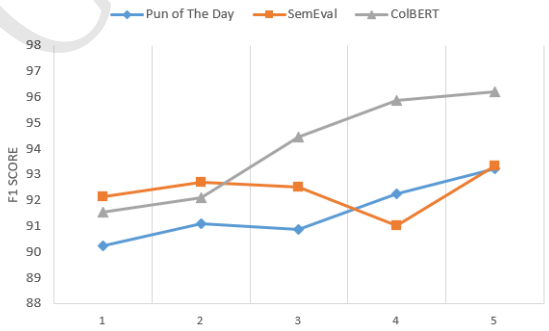

Figure 2: 不同数量的知识候选对模型性能的影响

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 Task 7-1a数据集而言，变化趋势于其他两个数据集不同。随着知识数量的增加，模型表现在略微提升之后，呈现出下降趋势，并且在$l = 4$时取得最差的结果，但是在$l = 5$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。

6 总结与展望

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

Issa Annamoradnejad and Gohar Zoghi. 2020. Colbert: Using bert sentence embedding for humor detection. *arXiv preprint arXiv:2004.12765*, 1(3).

Dario Bertero and Pascale Fung. 2016. A long short-term memory framework for predicting humor in dialogues. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 130–135.

Vladislav Blinov, Valeria Bolotova-Baranova, and Pavel Braslavski. 2019. Large dataset and language model fun-tuning for humor recognition. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 4027–4032.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. Comet: Commonsense transformers for automatic knowledge graph construction. *arXiv preprint arXiv:1906.05317*.

Andrew Cattle and Xiaojuan Ma. 2018. Recognizing humour using word associations and humour anchor extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1849–1858, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Dushyant Singh Chauhan, Gopendra Vikram Singh, Aseem Arora, Asif Ekbal, and Pushpak Bhattacharyya. 2022. A sentiment and emotion aware multimodal multiparty humor recognition in multilingual conversational setting. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6752–6761.

Peng-Yu Chen and Von-Wun Soo. 2018. Humor recognition using deep learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 113–117, New Orleans, Louisiana, June. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yufeng Diao, Liang Yang, Dongyu Zhang, Linhong Xu, Xiaochao Fan, Di Wu, and Hongfei Lin. 2018. Homographic puns recognition based on latent semantic structures. In *Natural Language Processing and Chinese Computing: 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8–12, 2017, Proceedings 6*, pages 565–576. Springer.

Xiaochao Fan, Hongfei Lin, Liang Yang, Yufeng Diao, Chen Shen, Yonghe Chu, and Yanbo Zou. 2020. Humor detection via an internal and external neural network. *Neurocomputing*, 394:105–111.

Alex Graves and Alex Graves. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45.

Md Kamrul Hasan, Wasifur Rahman, Amir Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, Louis-Philippe Morency, et al. 2019. Ur-funny: A multimodal language dataset for understanding humor. *arXiv preprint arXiv:1904.06618*.

Justine T. Kao, Roger Levy, and Noah D. Goodman. 2016. A computational model of linguistic humor in puns. *Cogn. Sci.*, 40(5):1270–1285.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521(7553):436–444.

Herbert M Lefcourt. 2001. *Humor: The psychology of living buoyantly*. Springer Science & Business Media.

Da Li, Rafal Rzepka, Michal Ptaszynski, and Kenji Araki. 2020. Hemos: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Information Processing & Management*, 57(6):102290.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018. Modeling sentiment association in discourse for humor recognition. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 586–591, Melbourne, Australia, July. Association for Computational Linguistics.

J. A. Meaney, Steven Wilson, Luis Chiruzzo, Adam Lopez, and Walid Magdy. 2021. SemEval 2021 task 7: HaHackathon, detecting and rating humor and offense. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 105–119. Association for Computational Linguistics, August.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538.

Rada Mihalcea, Carlo Strapparava, and Stephen Pulman. 2010. Computational models for incongruity detection in humour. In *Computational Linguistics and Intelligent Text Processing: 11th International Conference, CICLing 2010, Iaşi, Romania, March 21-27, 2010. Proceedings 11*, pages 364–374. Springer.

Tristan Miller and Iryna Gurevych. 2015. Automatic disambiguation of english puns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 719–729.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Alex Morales and Chengxiang Zhai. 2017. Identifying humor in reviews using background text sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 492–501, Copenhagen, Denmark, September. Association for Computational Linguistics.

Reynier Ortega-Bueno, Carlos E Muniz-Cuza, José E Medina Pagola, and Paolo Rosso. 2018. Uo upv: Deep linguistic humor detection in spanish social media. In *Proceedings of the third workshop on evaluation of human language technologies for Iberian languages (IberEval 2018) co-located with 34th conference of the Spanish society for natural language processing (SEPLN 2018)*, pages 204–213.

John Allen Paulos. 2008. *Mathematics and humor*. University of Chicago Press.

Victor Raskin. 1979. Semantic mechanisms of humor. In *Annual Meeting of the Berkeley Linguistics Society*, volume 5, pages 325–335.

Lu Ren, Bo Xu, Hongfei Lin, and Liang Yang. 2021. ABML: attention-based multi-task learning for jointly humor recognition and pun detection. *Soft Comput.*, 25(22):14109–14118.

Lu Ren, Bo Xu, Hongfei Lin, Jinhui Zhang, and Liang Yang. 2022. An attention network via pronunciation, lexicon and syntax for humor recognition. *Applied Intelligence*, 52(3):2690–2702.

Antonio Reyes and Paolo Rosso. 2012. Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision support systems*, 53(4):754–760.

Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.

Graeme Ritchie. 2009. Can computers create humor? *AI Magazine*, 30(3):71–71.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3027–3035.

Jerry M Suls. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *The psychology of humor: Theoretical perspectives and empirical issues*, 1:81–100.

Orion Weller and Kevin Seppi. 2019. Humor detection: A transformer gets the last laugh. *arXiv preprint arXiv:1909.00252*.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2367–2376.

Tongyue Zhang, Shaowu Zhang, Bo Xu, Liang Yang, and Hongfei Lin. 2021. 结合标签转移关系的多任务笑点识别方法(multi-task punchlines recognition method combined with label transfer relationship). In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 238–247, Huhhot, China, August. Chinese Information Processing Society of China.

樊小超, 杨亮, 林鸿飞, 刁宇峰, 申晨, 楚永贺, and 张桐. 2021. 基于多维潜在语义特征的幽默识别. 中文信息学报, 35(8):38–46.