·人工智能与模式识别·　文章编号：1000-3428(2020)08-0064-08　文献标志码：A　中图分类号：TP391

基于语言学特征与层次注意力机制的幽默识别

杨　勇[1a]，杨　亮[2]，邹艳波[1b]，任　鸽[1a]，樊小超[1a,2]

（1. 新疆师范大学 a. 计算机科学技术学院；b. 物理与电子工程学院，乌鲁木齐 830054；
2. 大连理工大学 计算机科学与技术学院，辽宁 大连 116024）

摘　要：结合英文幽默语言学特征，提出基于语音、字形和语义的层次注意力神经网络模型（PFSHAN）进行幽默识别。在特征提取阶段，将幽默文本表示为音素、字符以及携带歧义性等级信息的语义形式，分别采用卷积神经网络、双向门控循环单元和注意力机制提取 PFSHAN 模型的语音、字形和语义特征。在特征融合阶段，针对不同单词对幽默语言学特征的贡献程度不同，且不同幽默语言学特征和语句之间关联程度不同的问题，采用层次注意力机制调整不同幽默语言学特征对于 PFSHAN 模型性能的影响。在 Puns 和 Onliner 数据集上的实验结果表明，PFSHAN 模型的 F1 值分别为 91.03% 和 91.11%，能有效提高幽默识别性能。

关键词：幽默识别；注意力机制；卷积神经网络；双向门控循环单元；语言学特征

开放科学（资源服务）标志码（OSID）：

中文引用格式：杨勇，杨亮，邹艳波，等. 基于语言学特征与层次注意力机制的幽默识别[J]. 计算机工程，2020，46(8)：64-71.

英文引用格式：YANG Yong，YANG Liang，ZOU Yanbo，et al. Humor recognition based on linguistic features and hierarchical attention mechanism[J]. Computer Engineering，2020，46(8)：64-71.

Humor Recognition Based on Linguistic Features and Hierarchical Attention Mechanism

YANG Yong[1a]，YANG Liang[2]，ZOU Yanbo[1b]，REN Ge[1a]，FAN Xiaochao[1a,2]

（1a. School of Computer Science and Technology；
1b. School of Physics and Electronic Engineering，Xinjiang Normal University，Urumqi 830054，China；
2. School of Computer Science and Technology，Dalian University of Technology，Dalian，Liaoning 116024，China）

【Abstract】This paper proposes a hierarchical attention mechanism neural network model based on the features of pronunciation，font and semantics（PFSHAN） for humor recognition，which extracts the features of English humor linguistics. During the feature extraction stage，the humor texts are presented phoneme，character and semantic information that carries ambiguity level information，and then the features of pronunciation，font and semantics of the PFSHAN model are extracted by using the Convolutional Neural Network（CNN），Bi-directional Gated Recurrent Unit（Bi-GRU），and the attention mechanism. During the feature fusion stage，as words contribute differently to the linguistic features of humors，and the linguistic features of humors are also correlated differently to sentences，the hierarchical attention mechanism is used to adjust the influence of different linguistic features on performance of the PFSHAN model. Experimental results on datasets of Puns and Onliner show that the F1 scores of the PFSHAN model are 91.03% and 91.11% respectively，significantly improving the humor recognition performance.

【Key words】humor recognition；attention mechanism；Convolutional Neural Network（CNN）；Bi-directional Gated Recurrent Unit（Bi-GRU）；linguistic feature

DOI：10.19678/j.issn.1000-3428.0057138

基金项目：国家语委“十三五”科研规划项目（YB135-8）；新疆维吾尔自治区高等学校科研计划（XJEDU2016S066）；新疆师范大学博士科研启动基金（XJNUBS1609）。

作者简介：杨　勇（1979—），男，副教授、博士，主研方向为自然语言处理；杨　亮，讲师、博士；邹艳波，副教授、博士；任　鸽、樊小超（通信作者），讲师、硕士。

收稿日期：2020-01-06　**修回日期：**2020-03-20　**E-mail：**68523593@qq.com

计算机工程 www.ecice06.com

0　概述

幽默普遍存在于日常用语中,是人们沟通交流的重要组成部分。幽默一词来源于英文单词“Humor”,由林语堂先生于 1924 年引入中国,有可笑、有趣而意味深长之义[1]。近年来,随着人工智能的快速发展,幽默识别受到了国内外学者的广泛关注。幽默识别任务通常是识别某个语句或段落是否包含幽默的语义表达[2-3]。幽默数据集有多种类型[3],包括笑话、One-liner 形式的幽默、对话幽默等,本文的研究重点为 One-liner 形式的幽默。

One-liner 形式的幽默通常是一个简短的句子,使用少量词汇传达幽默的语义。与其他形式的幽默相比,One-liner 形式的幽默缺乏上下文信息,多数采用语音、语言歧义或叠字等手段产生预期的幽默效果。针对 One-liner 形式的幽默,目前的幽默识别方法主要分为基于特征工程的机器学习方法[4-5]和基于神经网络的深度学习方法[6-7]。前者需要领域专家构建特征,且耗时耗力,泛化能力较差。后者网络结构的构建通常缺乏幽默理论的驱动,可解释性较差。为解决以上问题,本文提出基于语音、字形和语义的层次注意力神经网络模型(PFSHAN)进行幽默识别。

1　相关工作

随着幽默在互联网中的广泛应用以及文本情感分析问题的深入研究,越来越多的学者对幽默识别产生了很大兴趣,幽默识别成为自然语言处理领域的热点研究问题之一。对于幽默识别研究,根据使用方法的不同,本文从基于特征工程的机器学习方法和基于神经网络的深度学习方法两个方面对现有工作进行概述。

基于特征工程的机器学习方法被广泛应用于幽默识别领域。文献[8]构建大规模的笑话语料库,并利用 n-gram 特征对幽默段落进行识别。文献[5]定义 3 种类型的幽默特征,包括头韵、反义词和成人俚语,并通过实验证明了其在幽默识别中的有效性。文献[9]基于幽默的不一致性理论和语言学特点,设计 5 个类别多达 50 多种幽默特征。文献[4]对幽默的潜在语义特征进行系统阐述并构建包括语音特征、歧义特征、不一致性特征和情感特征在内的 4 种类型的幽默特征。在此基础上,文献[10]将语义分析和情感分析相结合,对情感关联模式进行建模并用于幽默识别。文献[11]通过成分分析和依赖关系分析得到幽默的句法特征来提升幽默识别的性能。文献[12]基于幽默的歧义性和语音特性提出一系列幽默特征。文献[13]由喜剧电视节目中的对话构造了幽默数据集,并采用多模态的分析方法,结合声音特征与语义特征进行幽默识别。

近年来,基于神经网络的深度学习方法在幽默识别领域取得了许多研究成果。文献[14]提取《生活大爆炸》中的对话文本,利用幽默情景剧中特有的背景笑声自动标注笑点,并采用长短期记忆(Long Short Term Memory,LSTM)网络提取语义特征和声音特征识别笑点。文献[15]采用卷积神经网络(Convolutional Neural Network,CNN)和 LSTM 提取幽默特征并识别对话中的笑点。文献[7]比较 CNN 与传统机器学习方法的性能。文献[16]采用 LSTM 和注意力机制在幽默评测中取得了较好的结果。文献[17]结合人工特征和神经网络自动提取的特征,对西班牙语的推特文本进行幽默识别。文献[18]构建了一个大型的俄语幽默数据集,并使用调优的预训练语言模型进行幽默识别。文献[19]提出基于张量的幽默识别方法,能够有效提取幽默语句的词汇特征。

对于现有工作的研究结果表明,语音特征和歧义性特征能够有效提高幽默识别的性能,然而人工构造的特征成本较高且泛化能力较差。相比于基于特征工程的机器学习方法,基于神经网络的深度学习方法能够自动提取幽默的高维语义特征且性能较好。然而,现有基于神经网络的深度学习方法缺乏幽默理论的驱动,实验结果难以给出令人信服的解释。本文提出 PFSHAN 模型识别幽默语句,PFSHAN 模型基于幽默的语言学特征,分别从文本的语音、字形和语义 3 个维度提取幽默特征,并采用层次注意力机制,使得模型能够提取更有效的幽默特征。

2　基于音形义的幽默识别方法

如图 1 所示,本文提出基于音形义的层次注意力神经网络模型进行幽默识别的主要步骤为:1)将文本内容表示成对应的音素形式,采用卷积神经网络提取语句的语音特征;2)将文本表示成字符形式,采用双向门控循环单元(Bi-directional Gated Recurrent Unit,Bi-GRU)和注意力机制提取文本的字形特征;3)引入单词歧义性等级信息,更好地提取幽默语句的语义特征。为更好地区分不同幽默特征在幽默识别过程中的贡献程度,本文采用层级注意力机制来调节幽默语言学特征和幽默语句的关联程度。

计算机工程 www.ecice.com

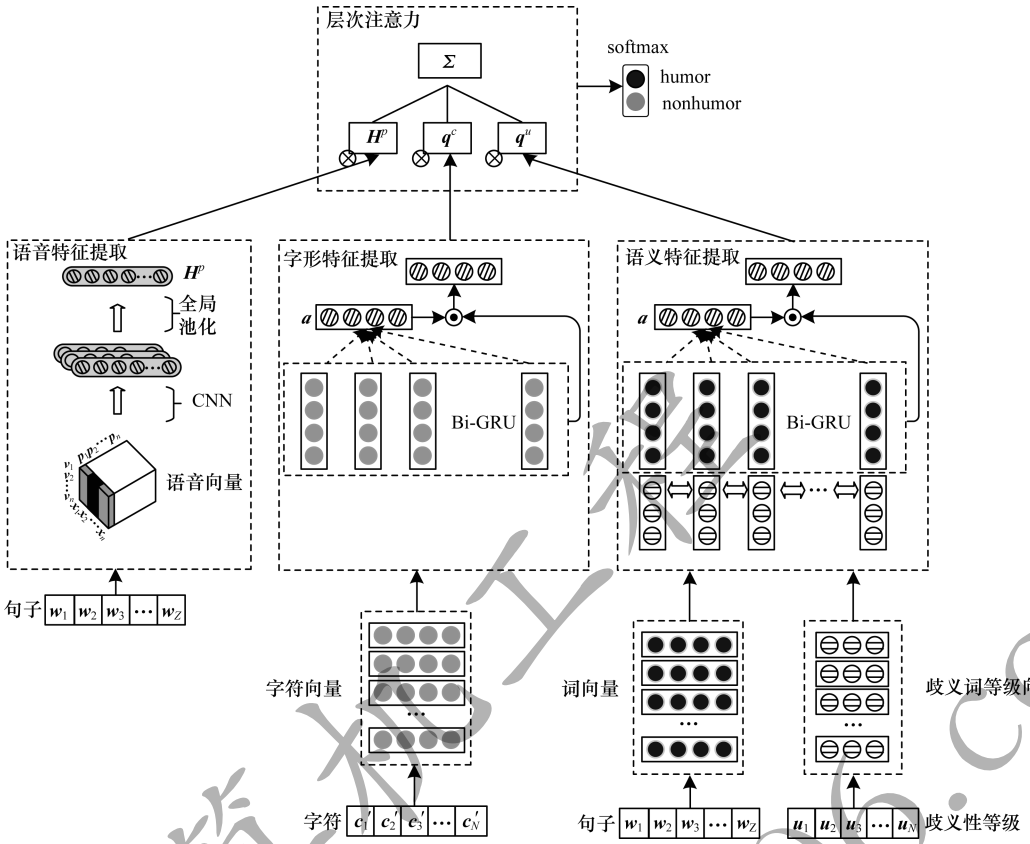

图 1 基于音形义的层次注意力神经网络模型

Fig. 1 Hierarchical attention neural network model based on pronunciation, font and semantics

2.1 基于语音的幽默特征提取

许多幽默由语音引起,文本内容中不协调的发音产生了幽默[20]。文献[5]指出幽默文本的语音特征与其语义内容一样重要。语音是引发幽默的重要手段,其通常通过押头韵或尾韵的形式进行表现[4]。

例 1 You can tune a piano, but you can't tuna fish.

在例 1 中,句子的语义并不有趣,但是句子中单词"tune"和"tuna"有相似的发音,这使得句子的幽默效果得到了加强。在许多幽默文本中,即使文本内容不幽默,也经常使用头韵、尾韵等语音特点引发或增强幽默效果。

由于单词的发音和拼写并不完全一致,因此无法从字符来直接获取句子的语音表示。为获得单词的语音表示,本文使用卡内基梅隆大学(CMU)的发音词典将文本表示成其对应的语音形式。相比于含有重音标识的版本,包含 39 个音素的无重音标识的 CMU 发音词典更加准确。因此,本文采用无重音标识的 CMU 发音词典将幽默语句中的单词转换成对应的音素表示。例如,单词"word"的音素表示为["W","ER","D"]。卷积神经网络能够更好地提取数据的局部特征且速度较快,因此本文采用卷积神经网络提取幽默语句中头韵、尾韵等语音特征。

1)语音的向量表示层。设语句 $\boldsymbol{S}=\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_N\}$,$\boldsymbol{w}_i$ 为语句中的任一单词,N 为句子长度,将 $\boldsymbol{w}_i$ 转换为对应的音素表示形式 $\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_L\}$,$\boldsymbol{p}_i\in\mathbb{R}^d$,$\boldsymbol{p}_i$ 为单词中任一音素的向量,d 为向量的维度,L 为某一单词的音素长度。对于长度小于 $L_{\max}$ 的单词和长度小于 $N_{\max}$ 的句子,使用零向量进行填充,并随机初始化语音向量。

2)变换层。本文的目标是发现单词间的头韵、尾韵等语音特征,因此采用变换层对输入张量进行变换,使得卷积神经网络的滑动窗口能够提取多个单词对应位置上的语音信息。

3)卷积层。卷积层利用一个窗口大小为 h 的卷积核提取局部的语音特征,其计算公式如下:

$$\boldsymbol{c}_i=f(\boldsymbol{w}\boldsymbol{p}_{i:i+h-1}+\boldsymbol{b}) \tag{1}$$

其中,$\boldsymbol{c}_i$ 为输出的特征向量,f 为非线性激活函数 ReLU,$\boldsymbol{w}$ 为参数,$\boldsymbol{p}_{i:i+h-1}$ 代表 $\boldsymbol{p}$ 中的第 i 列到第 $i+h-1$ 列,$\boldsymbol{b}$ 为偏置项。在实验中使用二维卷积神经网络及多个卷积核。

4)池化层。该层主要用于文本语音特征的降维,压缩参数数量,缓解过拟合现象,提高模型的容错能力。常用的池化操作有平均池化和最大池化两种策略,本文采用最大池化策略获取固定长度的语音特征向量:

$$\hat{\boldsymbol{C}}=\max(\boldsymbol{c}_i) \tag{2}$$

对池化后的特征向量进行拼接后,得到语句的

计算机工程 www.ceaj.com

语音特征表示为:

$$\boldsymbol{H}^{p}=[\hat{\boldsymbol{C}}_1,\hat{\boldsymbol{C}}_2,\cdots,\hat{\boldsymbol{C}}_n] \tag{3}$$

2.2　基于字形的幽默特征提取

幽默是一种文体,通常有其独特的表达方式,在很多情况下,正是字形的特征产生了幽默效果[21]。文献[22]指出反复出现的文本元素序列使得文本表现出相对稳定的特征。幽默语句常采用重复的字符或重复的标点符号等方法表达出幽默的效果。

例 2　I used to be a coyote, but I'm alright nooooooooooow!!!

例 2 是一个幽默的语句,该句采用字符重复的方式表现出幽默的效果。语句中的单词"now"是一个不规范的拼写形式,字符"o"被重复了多次,同时为了表达强调的效果,"!"也被重复了多次。这种刻意的字符重复是幽默语句的重要特征。

对于例 2 中"now"的不规范拼写形式,常规的词向量表示会将其作为未登录词处理,模型无法关注到该类单词对幽默识别性能的影响。为使模型能够捕获幽默语句的字形特征,本文对幽默语句的字符进行建模,将句子表示成字符的序列,句子的字符序列的向量表示作为模型输入。循环神经网络(Recurrent Neural Network, RNN)能够更好地处理序列信息,因此本文采用 RNN 提取语句中的重复字符、符号等字形特征。

在字形的向量表示层中,将目标语句中的每个单词 $\boldsymbol{w}_i$ 转换成对应的字符序列 $\{\boldsymbol{c}'_1,\boldsymbol{c}'_2,\cdots,\boldsymbol{c}'_m\}$,$\boldsymbol{c}'_i\in\mathbb{R}^{d''}$,$\boldsymbol{c}'_i$为单词中的任一字符的向量表示,$d''$为字符向量的维度,并且随机初始化字形向量。

在字形特征提取层中,为缓解 RNN 的梯度爆炸、梯度消失及长期依赖等问题,研究人员提出 LSTM 网络和门控循环单元(Gated Recurrent Unit, GRU)神经网络。GRU 相比 LSTM 参数更少,训练速度更快,而两者性能相当。基于以上特性,本文采用 GRU 提取字形特征。GRU 利用重置门和更新门控制序列的状态更新。在 t 时刻 GRU 的状态可以形式化表示为:

$$\boldsymbol{z}_t=\sigma(\boldsymbol{W}_z\boldsymbol{x}_t+\boldsymbol{U}_z\boldsymbol{h}_{t-1}+\boldsymbol{b}_z) \tag{4}$$

$$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_r\boldsymbol{x}_t+\boldsymbol{U}_r\boldsymbol{h}_{t-1}+\boldsymbol{b}_r) \tag{5}$$

$$\tilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_h\boldsymbol{x}_t+\boldsymbol{r}_t\odot\boldsymbol{U}_h\boldsymbol{h}_{t-1}+\boldsymbol{b}_n) \tag{6}$$

$$\boldsymbol{h}_t=\boldsymbol{z}_t\boldsymbol{h}_{t-1}+(1-\boldsymbol{z}_t)\odot\tilde{\boldsymbol{h}}_t \tag{7}$$

其中,$\boldsymbol{z}_t$ 为更新门,$\boldsymbol{r}_t$ 为重置门,$\boldsymbol{h}_t$ 为 t 时刻的隐层状态,$\tilde{\boldsymbol{h}}_t$ 为候选状态,$\boldsymbol{x}_t$ 为 t 时刻的输入,σ 为 sigmoid 函数,tanh 为激活函数,$\boldsymbol{W}_z$、$\boldsymbol{W}_r$、$\boldsymbol{W}_h$、$\boldsymbol{U}_z$、$\boldsymbol{U}_r$、$\boldsymbol{U}_h$ 为参数,$\boldsymbol{b}_z$、$\boldsymbol{b}_r$、$\boldsymbol{b}_h$ 为偏置项,$\odot$为对应元素相乘。

GRU 能够提取每个时间步长 t 之前的信息,但是忽略了 t 之后的文本信息。Bi-GRU 包含两个相互独立的隐藏状态,可以同时从前向和后向提取文本信息,然后对两部分信息进行整合,从而更好地利用文本的上下文信息。本文采用 Bi-GRU 提取文本的字形特征,其形式化表示如下:

$$\overrightarrow{\boldsymbol{h}}_t=\mathrm{GRU}(\overrightarrow{\boldsymbol{h}}_t,e(\boldsymbol{w}_t)) \tag{8}$$

$$\overleftarrow{\boldsymbol{h}}_t=\mathrm{GRU}(\overleftarrow{\boldsymbol{h}}_t,e(\boldsymbol{w}_t)) \tag{9}$$

$$\boldsymbol{h}_t=[\overrightarrow{\boldsymbol{h}}_s\cdot\overleftarrow{\boldsymbol{h}}_s] \tag{10}$$

其中,$\overrightarrow{\boldsymbol{h}}_t$ 表示前向 GRU 的隐藏状态,$\overleftarrow{\boldsymbol{h}}_t$ 表示后向 GRU 隐藏状态,$\boldsymbol{h}_t$ 为两个方向隐藏状态的拼接结果。字形特征可表示为 $\boldsymbol{H}^c=\text{Bi-GRU}(\boldsymbol{c}'_t,\boldsymbol{h}_{t-1})$。

在字符特征注意力层中,为能够对携带显著语义信息的字符给予更多的关注,在提取字形特征时,引入注意力机制,其形式化表示如下:

$$\boldsymbol{w}_{ij}=\tanh(\boldsymbol{W}^{\mathrm{T}}[\boldsymbol{h}_j\cdot\boldsymbol{H}^c]+\boldsymbol{b}) \tag{11}$$

$$\boldsymbol{a}_{ij}=\frac{\exp(\boldsymbol{w}_{ij})}{\sum_{j=1}^{n}\exp(\boldsymbol{w}_{ij})} \tag{12}$$

$$\boldsymbol{q}^c=\sum_{j=1}^{n}\boldsymbol{a}_{ij}\boldsymbol{h}_j \tag{13}$$

其中,$\boldsymbol{W}$ 为权重矩阵,$\boldsymbol{b}$ 为偏置项,tanh 为激活函数,$\boldsymbol{a}_{ij}$为注意力权重,所有参数采用随机初始化并在训练中动态更新,$\boldsymbol{q}^c$ 为字符特征注意力层的输出向量。

2.3　基于语义的幽默特征提取

句子本身的语义特征将为幽默识别提供直接的线索。文献[23]指出语义的歧义性会引发幽默,歧义性是幽默产生的重要因素。幽默语句中的歧义性是指句子中的某些单词包含多个语义,使得句子存在多种不同的理解方式[24]。

例 3　Did you hear about the guy whose whole left side was cut off? He's all right now.

例 3 是一个典型的由于歧义性引起幽默的语句。单词"right"包含多个语义,它既可以被理解为"右侧",又可以被理解为"恢复"。由于单词的多个语义造成了句子理解的偏差,因此使该语句显得十分有趣。句子中单词包含的同义词的个数与语句是否幽默具有一定的相关性。

基于特征工程的机器学习方法将单词包含的同义词的个数作为特征来识别幽默[4]。为使神经网络模型能够学习到包含不同同义词数量的单词,本文根据同义词的个数对单词进行分类,将类别信息进行向量表示并和单词的向量表示进行融合,最后采用 Bi-GRU 和注意力机制提取携带歧义性信息的潜在语义特征。

在词向量表示层中,基于分布式假设将单词映射为低维稠密向量,同时保持单词的语义信息。设语句的向量表示为 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$,$\boldsymbol{x}_i\in\mathbb{R}^d$,$\boldsymbol{x}_i$ 为词向量表示,d 为词向量维度。本文采用 GloVe[25] 作为预训练词向量对词向量进行初始化。

在歧义性等级向量表示层中,对于给定的语句 $\boldsymbol{S}=\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_N\}$,采用 WordNet 计算每个单词 $\boldsymbol{w}_i$ 的同义词数量,根据同义词数量的分布情况将单词

计算机工程 www.ecice.com

划分成多个类别。由于停用词对幽默语句的贡献较小,因此停用词被单独划分为一个类别。将每个单词表示成其对应的歧义性类别并采用随机初始化的方式得到歧义性等级的向量表示 $\boldsymbol{U}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_N)$,$\boldsymbol{u}_i\in\mathbb{R}^k$ 表示每个单词的歧义性等级向量,k 为向量维度,N 为句子长度。

在特征融合层中,为使模型能够更好地学习歧义性信息和语义信息,本文将歧义性等级信息和语义信息进行融合 $\boldsymbol{x}_i'=\boldsymbol{x}_i\oplus\boldsymbol{u}_i,\boldsymbol{x}_i'\in\mathbb{R}^{d+k}$。

在语义特征提取层中,Bi-GRU 能够有效处理文本序列数据并能够更好地提取上下文信息。因此,本文采用 Bi-GRU 提取文本的语义特征,携带歧义性等级信息的语义特征可表示为 $\boldsymbol{H}^u=\text{Bi-GRU}(\boldsymbol{x}_i',\boldsymbol{h}_{i-1})$。

在语义特征注意力层中,为使模型能够关注携带显著语义信息的单词,在提取语义特征时,引入注意力机制,其中 $\boldsymbol{q}^u$ 为语义特征注意力层的输出向量。

2.4 层次注意力机制

由于不同幽默语言学特征和幽默语句的关联程度不同,因此本文采用层次注意力机制调整不同语言学特征对于幽默识别性能的影响,其形式化表示如下:

$$\boldsymbol{w}_j=\tanh(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{V}_j+\boldsymbol{b}) \tag{14}$$

$$\boldsymbol{\beta}_j=\frac{\exp(\boldsymbol{w}_j)}{\sum_{j=1}^{5}\exp(\boldsymbol{w}_j)} \tag{15}$$

$$\boldsymbol{q}=\sum_{j=1}^{5}\boldsymbol{\beta}_j\boldsymbol{k}_j\boldsymbol{R}_j\in\{\boldsymbol{H}^p,\boldsymbol{q}^c,\boldsymbol{q}^u\} \tag{16}$$

其中,$\boldsymbol{W}$ 为权重矩阵,$\boldsymbol{b}$ 为偏置项,$\boldsymbol{H}^p$ 为语音特征表示,$\boldsymbol{q}^c$ 为字形特征表示,$\boldsymbol{q}^u$ 为语义特征表示,$\boldsymbol{V}_j$ 为不同句子的表示,$\boldsymbol{\beta}_j$ 为注意力权重,所有参数采用随机初始化并在训练中动态更新,$\boldsymbol{q}$ 为句子的最终特征表示。

2.5 幽默分类

本文提取文本的语音、字形和语义特征,采用 softmax 函数进行幽默识别,其形式化表示如下:

$$\boldsymbol{v}=\tanh(\boldsymbol{W}_p\boldsymbol{q}+\boldsymbol{b}_p) \tag{17}$$

$$\hat{\boldsymbol{y}}=\text{softmax}(\boldsymbol{W}_f\boldsymbol{v}+\boldsymbol{b}_f) \tag{18}$$

其中,$\boldsymbol{W}_p$、$\boldsymbol{W}_f$ 为参数矩阵,$\boldsymbol{b}_p$、$\boldsymbol{b}_f$ 为偏置项,$\hat{\boldsymbol{y}}$ 为预测标签。

本文模型基于反向传播算法与端到端的方式进行训练,并采用期望交叉熵作为损失函数。

$$\text{loss}=-\sum_i\sum_j\boldsymbol{y}_i^j\log_a\hat{\boldsymbol{y}}_i^j+\lambda\|\theta\|^2 \tag{19}$$

其中,$\boldsymbol{y}$ 为真实标签,i、j 分别为句子的编号和类别编号,λ 为正则化参数,θ 为超参数。

3 实验结果与分析

3.1 实验数据与评价指标

Puns 数据集[4]中的幽默语句来自同名网站,非幽默文本来自美联社新闻、纽约时报、雅虎新闻和谚语。Puns 数据集包含幽默语句 2 423 条,非幽默语句 2 403 条,句子平均长度为 13.5。Oliner 数据集[5]中的幽默语句来自多个著名的幽默网站,非幽默语句来自路透社新闻标题。Oliner 包含幽默、非幽默语句各 16 000 条,句子平均长度为 12.6。为便于和基线方法进行比较,本文采用精确率、准确率、查全率和 F1 值作为评价指标。

3.2 实验设置

在训练过程中,词向量采用 GloVe 进行初始化,维度为 300。语音向量采用高斯分布 $\boldsymbol{U}(-0.1,0.1)$ 进行随机初始化,维度为 100。字符向量采用随机初始化,维度为 100。单词被划分为 4 个歧义性类别,歧义性等级采用随机初始化,维度为 10。卷积神经网络采用 2D 卷积和池化层,卷积核数量为 128,卷积核大小为 2、3、4。Bi-GRU 的神经元个数为 150,优化方法为 Adadelta[26]。Batch 大小为 64,dropout 为 0.5。同时,在训练过程中使用学习率衰减和早停机制防止过度拟合,并使用五倍交叉验证法减少数据集划分的影响。

3.3 对比方法

实验对比方法具体如下:

1)支持向量机(Support Vector Machine,SVM)。该方法[4]使用人工构造的语音特征、歧义特征、不一致特征和情感特征,采用支持向量机模型。

2)HCFW2V。该方法[4]同时使用上述 4 类特征和词向量作为幽默特征,采用随机森林模型。

3)ST。该方法[10]同时使用上述 4 类特征以及人工构造的情感冲突和情感转换特征,采用随机森林模型。

4)Syn。该模型[11]同时使用上述 4 类特征以及人工构造的句法结构特征,采用随机森林模型。

5)CNN。该模型[7]采用卷积神经网络进行幽默识别。

6)Bi-GRU。该模型采用 Bi-GRU 提取幽默文本的潜在语义特征并进行幽默识别。

7)Bi-GRU + Att。该模型采用 Bi-GRU 和注意力机制提取语义特征并进行幽默识别。

8)CNN + HN。该模型[27]采用 CNN 和 Highway 网络架构。

9)PFSHAN。本文提出的一种基于语音、字形和语义的层次注意力神经网络模型。

表 1 和表 2 列出了不同幽默识别方法与模型的性能对比,其中最佳结果加粗显示,实验结果表明:

1)基于特征工程的机器学习方法的性能低于基于神经网络的深度学习方法。对于相同的人工特征集合,基于特征工程的机器学习方法在两个数据集上性能有所差别。HCFW2V 在 Puns 数据集上性能较好,而 SVM 在 Oliner 数据集上性能较好。这也说

www.ecice06.com 计算机工程

明了基于特征工程的机器学习方法依赖于人工特征的构造,其泛化能力较差。此外,引入句法信息后,幽默识别的性能有了一定幅度的提升。

2)基于神经网络的深度学习方法能够自动学习幽默语句的潜在语义特征,在两个数据集上均表现出较好的性能。Bi-GRU 能够更好地利用上下文信息与长距离的依赖关系,其性能优于 CNN。引入 Highway 后,CNN 的性能有了较大幅度的提升。

3)PFSHAN 模型在两个数据集上均取得了最佳的性能。PFSHAN 模型能够提取语句的语音、字形和语义信息,而且在提取语义特征时,其能够捕获单词的歧义性信息,从多个维度提取幽默特征。此外,PFSHAN 模型采用层级注意力机制,不仅能够调节不同输入对提取特征的影响,而且能够调节不同语言学特征对幽默识别的影响。

表 1　Puns 数据集上的实验结果

Table 1　Experimental results on Puns dataset　%

方法与模型	精确率	准确率	查全率	F1 值
SVM	83.12	88.04	80.26	82.24
HCFW2V	85.40	83.40	88.80	85.90
CNN	86.10	86.40	86.40	85.70
CNN + HW	89.40	86.60	**94.00**	90.10
Bi-GRU	87.51	84.21	91.76	87.82
Bi-GRU + Att	87.76	85.42	91.88	88.53
PFSHAN	**90.12**	**89.68**	92.43	**91.03**

表 2　Oliner 数据集上的实验结果

Table 2　Experimental results on Oliner datasets　%

方法与模型	精确率	准确率	查全率	F1 值
SVM	83.85	85.91	82.52	84.18
HCFW2V	79.70	77.60	83.60	80.50
ST	81.30	81.40	82.30	81.80
Syn	85.00	82.70	89.10	85.80
CNN	84.24	85.73	86.46	86.09
CNN + HW	**89.70**	87.20	**93.60**	90.30
Bi-GRU	85.69	87.62	84.82	86.20
Bi-GRU + Att	85.88	88.10	84.97	86.51
PFSHAN	89.68	**89.43**	92.86	**91.11**

3.4　歧义性等级信息对模型性能的影响

为验证歧义性等级信息对幽默识别的影响,本文对比仅使用语义信息的 Bi-GRU 和加入歧义性等级信息的 Bi-GRU 的 PFSHAN 模型幽默识别性能。如图 2 所示,加入了歧义性等级信息后,PFSHAN 模型 F1 值均有所提高,在 Puns 数据集上 F1 值提高了 0.8%,在 Oliner 数据集上提高了 1.14%。实验结果表明,单词的歧义性等级信息能够有效提高 PFSHAN 模型的幽默识别性能。

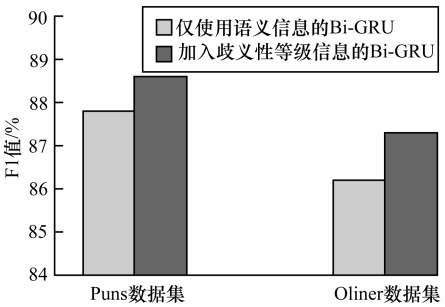

图 2　歧义性等级信息对幽默识别性能的影响

Fig. 2　Impact of ambiguous level information on performance of humor recognition

3.5　语音、字形和语义特征对模型性能的影响

本文对比语音、字形和语义特征对 PFSHAN 模型性能的影响,PFSHAN-pro、PFSHAN-font、PFSHAN-sem 分别表示未使用语音、字形和语义信息的 PFSHAN 模型。如表 3 所示,当 PFSHAN 模型未使用语义信息时,模型性能受到的影响最大。这表明模型能够从文本的潜在语义信息中学习到与幽默关联较强的信息,如果仅从语音和字形特征对幽默进行识别,则模型性能较差。当 PFSHAN 模型未使用字形信息时,对模型性能影响较小。这可能是因为在构造数据时对数据进行了预处理,其不规范的拼写等字形特征较少。语音特征对模型有一定的影响,说明文本中一部分幽默是由语音特征引起。当同时引入音形义特征时,PFSHAN 模型取得了最佳的性能,这表明语音、字形和语义特征能够更加有效地对幽默文本进行表征,从而提高幽默识别性能。

表 3　语音、字形和语义特征对幽默识别性能的影响

Table 3　Impact of pronunciation, font and semantics on performance of humor recognition　%

模型	F1 值	
	Puns 数据集	Oliner 数据集
PFSHAN	91.03	91.11
PFSHAN-pro	89.93	90.01
PFSHAN-font	90.14	90.22
PFSHAN-sem	86.35	86.42

3.6　层次注意力机制对模型性能的影响

本文对比了不同注意力机制对幽默识别性能的影响。PFSHAN-Hyp 表示提取字形和语义特征后,采用注意力机制得到字形和语义信息的表示,然后直接和语音信息进行拼接并识别幽默。PFSHAN-Lin-Hyp 表示只使用 Bi-GRU 提取字形和语义特征,并使用 CNN 提取语音特征,然后拼接 3 类特征进行幽默识别。

计算机工程 www.ecocen.com

如表 4 所示,采用层次注意力机制能够有效提高幽默识别的性能,相比不使用注意力机制的模型,PFSHAN 在两个数据集上的 F1 值分别提高了 1.19% 和 0.97% 。实验结果表明,层次注意力机制不但能够调整不同字符或单词对于不同幽默特征的权重,而且能够调节不同幽默语言学特征和幽默语句的关联程度,从而提高幽默识别性能。

表 4 层次注意力机制对幽默识别性能的影响

Table 4 Impact of hierarchical attention mechanism on performance of humor recognition %

模型	F1 值	
	Puns 数据集	Oliner 数据集
PFSHAN	91.03	91.11
PFSHAN-Hyp	90.46	90.52
PFSHAN-Lin-Hyp	89.84	90.14

3.7 错误样例分析

为更好地研究并提升 PFSHAN 模型在幽默识别任务中的性能,对其错误样例进行分析。以下是两个 PFSHAN 模型不能正确识别的样例:

例 4 The one who invented the door knocker got a no bell prize.

例 5 A clean house is a sure sign of a broken computer.

例 4 和例 5 均为幽默样例,但是 PFSHAN 模型却把它们视为非幽默的语句。在例 4 中,“no bell prize”的发音和“nobel prize”发音十分类似,所以引发了幽默的效果。显然,该句的幽默效果是语音所致,但是“nobel prize”没有出现在原文中,PFSHAN 模型无法捕获相关的语音特征。此外,背景知识也是判断该语句是否是幽默的重要因素。在例 5 中,“clean house”和“broken computer”形成了语义上的对比,这种不协调、不一致使得句子产生了幽默的效果,因此如何捕获文本语义的不一致性将是未来幽默识别中的重要研究方向。

4 结束语

本文提出基于语音、字形和语义的层次注意力神经网络模型(PFSHAN)进行幽默识别。基于幽默文本的语言学特点,采用 CNN 和 Bi-GRU 捕获幽默语句的语音、字符和语义特征,同时利用层次注意力机制调节不同语言学特征对幽默识别的影响。实验结果表明,本文方法能够有效获取幽默语句的音形义特征,提高幽默识别性能。但由于 PFSHAN 模型仅适用于英文文本的幽默识别,而中英文表达在很多方面存在差异,因此下一步将构建中文幽默数据集及模型进行中文幽默文本识别。此外,如何利用自注意力机制与预训练模型捕获文本语义的不一致特征也将是今后研究的重点。

参考文献

[1] LIN Yutang. Humor life[M]. Xi'an: Shaanxi Normal University Press, 2002. (in Chinese)
林语堂. 幽默人生[M]. 西安:陕西师范大学出版社, 2002.

[2] LIN Hongfei, ZHANG Dongyu, YANG Liang, et al. Computational humor researches and applications[J]. Journal of Shandong University (Natural Science), 2016, 51(7): 1-10. (in Chinese)
林鸿飞,张冬瑜,杨亮,等. 幽默计算及其应用研究[J]. 山东大学学报(理学版), 2016, 51(7): 1-10.

[3] ATTARDO S. Linguistic theories of humor[J]. Language, 1996, 72: 45-64.

[4] YANG D, LAVIE A, DYER C, et al. Humor recognition and humor anchor extraction[C]//Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing. Washington D. C., USA: IEEE Press, 2015: 2367-2376.

[5] MIHALCEA R, STRAPPARAVA C. Making computers laugh: investigations in automatic humor recognition[C]// Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2005: 531-538.

[6] FAN Xiaochao, LIN Hongfei, YANG Liang, et al. Humor detection via an internal and external neural network[J]. Neurocomputing, 2020, 394: 105-111.

[7] CHEN L, CHONG M L. Predicting audience's laughter using convolutional neural network[EB/OL]. [2019-12-14]. https://arxiv.org/abs/1702.02584v2.

[8] REN Lu, YANG Liang, XU Linhong, et al. Construction and application of Chinese joke corpus[J]. Journal of Chinese Information Processing, 2018, 32(7): 20-29. (in Chinese)
任璐,杨亮,徐琳宏,等. 中文笑话语料库的构建与应用[J]. 中文信息学报, 2018, 32(7): 20-29.

[9] ZHANG Ren, LIU Neng. Recognizing humor on twitter[C]// Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2014: 889-898.

[10] LIU Lin, ZHANG Dong, SONG Wen. Modeling sentiment association in discourse for humor recognition[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018: 586-591.

[11] LIU Lin, ZHANG Dong, SONG Wen. Exploiting syntactic structures for humor recognition[C]//Proceedings of the 27th International Conference on Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018: 1875-1883.

[12] BARBIERI F, SAGGION H. Automatic detection of irony and humor in Twitter[C]//Proceedings of the 15th International Conference on Computational Creativity. Washington D.C., USA: IEEE Press, 2014: 155-162.

[13] PURANDARE A, LITMAN D. Humor: prosody analysis and automatic recognition for friends[C]//Proceedings of 2006 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2006: 208-215.

计算机工程 www.ecice06.com

[14] BERTERO D, FUNG P. A long short-term memory framework for predicting humor in dialogues[C]// Proceedings of 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Philadelphia, USA: Association for Computational Linguistics, 2016: 130-135.

[15] BERTERO D, FUNG P. Deep learning of audio and language features for humor prediction[C]//Proceedings of the 10th International Conference on Language Resources and Evaluation. Amsterdam, the Netherlands: [s. n.], 2016: 496-501.

[16] BAZIOTZIS C, PELEKIS N, DOULKERIDIS C. DataStories at SemeVal-2017 task 6: siamese LSTM with attention for humorous text comparison[C]// Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada: [s. n.], 2017: 390-395.

[17] ORTEGA-BUENO R, MUANIZ-CUZA C E, PAGOLA J E M, et al. UO UPV: deep linguistic humor detection in Spanish social media[C]//Proceedings of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages Co-located with the 34th Conference of the Spanish Society for Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2018: 203-213.

[18] BLINOV V, BOLOTOVA-BARANOVA V, BRASLAVSKI P. Large dataset and language model fun-tuning for humor recognition[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2019: 4027-4032.

[19] ZHAO Zi, CATTLE A, PAPALEXAKIS E, et al. Embedding lexical features via tensor decomposition for small sample humor recognition[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2019: 6377-6382.

[20] STOCK O, STRAPPARAVA C, NIJHOLT. The April fools' day workshop on computational humour[EB/OL]. [2019-12-14]. https://www.researchgate.net/publication/242716918.

[21] CASTRO S, CUBERO M, GARAT D, et al. Is this a joke? Detecting humor in Spanish Tweets[C]// Proceedings of Ibero-American Conference on Artificial Intelligence. Berlin, Germany: Springer, 2016: 139-150.

[22] REYES A, ROSSO P, BUSCALDI D. From humor recognition to irony detection: the figurative language of social media[J]. Data and Knowledge Engineering, 2012, 74: 1-12.

[23] REYES A, ROSSO P, VEALE T. A multidimensional approach for detecting irony in Twitter[J]. Language Resources and Evaluation, 2013, 47(1): 239-268.

[24] BUCARIA C. Lexical and syntactic ambiguity as a source of humor: the case of newspaper headlines[J]. Humor, 2004, 17(3): 279-310.

[25] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2014: 1532-1543.

[26] ZHANG R, GONG W G, GRZEDA V, et al. An adaptive learning rate method for improving adaptability of background models[J]. IEEE Signal Processing Letters, 2013, 20(12): 1266-1269.

[27] CHEN P Y, SOO V W. Humor recognition using deep learning[C]//Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Philadelphia, USA: Association for Computational Linguistics, 2018: 113-117.

编辑　陆燕菲

计算机工程 www.ceco.com