

# 基于对抗式神经网络的多维度情绪回归\*

朱苏阳, 李寿山, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215008)

通讯作者: 朱苏阳, E-mail: syzhu@stu.suda.edu.cn; 李寿山, E-mail: lishoushan@suda.edu.cn



**摘要:** 情绪分析是细粒度的情感分析任务,其目的是通过训练机器学习模型来判别文本中蕴含了何种情绪,是当前自然语言处理领域中的研究热点.情绪分析可细分为情绪分类与情绪回归两个任务.针对情绪回归任务,提出一种基于对抗式神经网络的多维度情绪回归方法.所提出的对抗式神经网络由3部分组成:特征抽取器、回归器、判别器.该方法旨在训练多个特征抽取器和回归器,以对输入文本的不同情绪维度进行打分.特征抽取器接受文本为输入,从文本中抽取针对不同情绪维度的特征;回归器接受由特征抽取器输出的特征为输入,对文本的不同情绪维度打分;判别器接受由特征抽取器输出的特征为输入,以判别输入的特征是针对何情绪维度.该方法借助判别器对不同的特征抽取器进行对抗式训练,从而获得能够抽取泛化性更强的针对不同情绪维度的特征抽取器.在EMOBANK多维度情绪回归语料上的实验结果表明,该方法在EMOBANK新闻领域和小说领域的情绪回归上均取得了较为显著的性能提升,并在 $r$ 值上超过了所有的基准系统,其中包括文本回归领域的先进系统.

**关键词:** 情绪回归;情绪分析;对抗式神经网络;对抗式学习;EMOBANK语料

**中图法分类号:** TP391

中文引用格式: 朱苏阳,李寿山,周国栋.基于对抗式神经网络的多维度情绪回归.软件学报,2019,30(7):2091–2108. <http://www.jos.org.cn/1000-9825/5838.htm>

英文引用格式: Zhu SY, Li SS, Zhou GD. Multi-dimensional emotion regression via adversarial neural network. Ruan Jian Xue Bao/Journal of Software, 2019,30(7):2091–2108 (in Chinese). <http://www.jos.org.cn/1000-9825/5838.htm>

## Multi-dimensional Emotion Regression via Adversarial Neural Network

ZHU Su-Yang, LI Shou-Shan, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215008, China)

**Abstract:** Emotion analysis, which aims to determine the emotion contained in a piece of text via training a machine learning model, is a fine-grained sentiment analysis task. Emotion analysis can be divided into two tasks: Emotion classification and emotion regression. In this paper, an adversarial neural network is proposed for multi-dimensional emotion regression task. The proposed network consists of three modules: Feature extractors, regressors, and a discriminator. The network aims to train multiple feature extractors and regressors to score for multiple emotion dimensions for a textual input. Feature extractors take a text as inputs, and extract different feature vectors for different emotion dimensions. Regressors take extracted feature vectors as inputs to score for multiple emotion dimensions. The discriminator take an extracted feature vector as its input, and discriminate for which emotion dimension the feature vector is extracted. The proposed approach conducts adversarial training between different feature extractors via the discriminator in order to training feature extractors which can extract more generalized features for multiple emotion dimensions. Empirical studies on EMOBANK corpus demonstrate the notable improvements in  $r$ -value achieved by the proposed approach on EMOBANK readers' and writers' emotion regression in news domain and fictions domain compared to all baseline systems, including several state-of-the-art text regression systems.

**Key words:** emotion regression; emotion analysis; adversarial neural network; adversarial learning; EMOBANK corpus

\* 基金项目: 国家自然科学基金(61672366, 61751206)

Foundation item: National Natural Science Foundation of China (61672366, 61751206)

收稿时间: 2018-05-27; 修改时间: 2019-01-07; 采用时间: 2019-03-26

情绪分析(emotion analysis)作为一种细粒度的情感分析(sentiment analysis)任务,旨在判别自然语言文本中所蕴含的情绪,是自然语言处理领域中的研究热点<sup>[1-3]</sup>。一条文本中通常包含两类情绪:作者情绪(作者以什么样的情绪写下这条文本)和读者情绪(读者看到这条文本会产生什么样的情绪)。而情绪分析可进一步分为以下两个任务。

1) 情绪分类(emotion classification).通过分类模型判别输入文本所属的情绪类别(emotion category)。例如,例句 1(取自 Yao 等人<sup>[2]</sup>的作者情绪分类语料库)表达了一种开心(happiness)的作者情绪。

2) 情绪回归(emotion regression).通过回归模型对输入文本的情绪维度(emotion dimension)进行打分。其中,情绪维度一般是多维的,包括极性(valence,简称 V)、强度(arousal,简称 A)以及可控性(dominance,简称 D)。例如,例句 2(取自 Buechel 等人<sup>[4]</sup>的多维度情绪回归语料库 EMOBANK)的 3 个作者情绪维度分数(V,A,D)=(2.4,3.6,2.8)。

例句 1:时间真的过的好快,我从一个未成年到奔三了,感谢一路上很多人,感谢我认识你们大家,谢谢你们。

例句 2:You see, Nathan, there is ...separate from your society ...oh damn it.(你懂的,内森,有这样的人...与你的社会所隔绝...哦,该死的。)

情绪分类的研究工作开始得较早,有许多相关工作,包括传统的基于统计模型的方法和最新的基于神经网络模型的方法。例如, Yang 等人<sup>[5]</sup>利用情感词典作为外部语言资源训练 SVM 分类器来判别博客文本所属的情绪类别。Tripathi 等人<sup>[6]</sup>利用深度卷积神经网络判别文本的情绪类别。然而,由于缺乏公认的情绪分类体系(emotion taxonomy),不同的情绪分类语料库大多会采用不同的情绪分类体系,其分类体系中包含的情绪类别也不尽相同。例如,例句 1 所在语料库使用的是基于 Ekman<sup>[7]</sup>的 6 类情绪分类体系之上设计的 7 类情绪分类体系,而例句 1 若在 Huang 等人<sup>[8]</sup>所使用的情绪体系下则会被标注为正面复合(positive complex)的作者情绪。因此,基于单一语料库的情绪分类任务所训练的分类模型很难简单复用到使用其他情绪分类体系的语料库上<sup>[9]</sup>。

与情绪分类相比,情绪回归任务受限于回归任务的难度和情绪回归语料库的缺乏,相关研究起步较晚。情绪回归语料库相比于情绪分类语料库的一大优势在于,其多使用公认的在心理学上用于描述情绪的极性-强度模型(valence-arousal model)<sup>[10]</sup>对语料进行情绪维度分数的标注。Buechel 等人<sup>[4]</sup>在极性-强度模型的基础上将其拓展为极性-强度-可控性模型(valence-arousal-dominance model),并基于这一模型构建了一个包含超过 10 000 条样本的多维度情绪回归语料库 EMOBANK。本文致力于在该语料库上进行多维度的情绪回归研究。

近年来,基于神经网络的方法在许多自然语言处理任务中都取得了较好的成果。例如,句法分析<sup>[11,12]</sup>、情感分析<sup>[13,14]</sup>、信息抽取<sup>[15,16]</sup>、机器翻译<sup>[17,18]</sup>等。在神经网络的相关研究中,生成式对抗网络(generative adversarial network,简称 GAN)通过对抗式学习(adversarial learning)来训练能够生成接近真实样本的生成器<sup>[19]</sup>。基于对抗式学习的学习框架在图像处理领域中已经取得了很好的成果<sup>[20,21]</sup>,在自然语言处理领域也被初步应用到文本生成任务中去<sup>[22,23]</sup>,这是一种很有潜力的学习框架。并且,该学习框架可被应用到全监督、半监督以及无监督的机器学习任务中。

目前,将对抗式学习框架应用于分类任务的相关研究还较少<sup>[24,25]</sup>,而应用于回归任务的研究则更为罕见。本文提出一种基于对抗式神经网络的多维度情绪回归方法。本文提出的方法包括对抗式神经网络(adversarial neural network)模型和对抗学习算法。对抗式神经网络包含特征抽取器、回归器、判别器这 3 个部分,通过联合训练多个特征抽取器和回归器,以对输入文本的不同情绪维度进行打分。而对抗学习算法借由判别器在不同的特征抽取器之间进行对抗式训练,从而获得能够抽取出泛化性更强的针对不同情绪维度的特征抽取器,以提高情绪回归的性能。在 EMOBANK 多维度情绪回归语料库上的实验结果表明,本文的方法优于其他基准系统和领域内先进系统,在 EMOBANK 新闻领域和小说领域的情绪回归上均取得了较好的性能。

本文第 1 节介绍情绪分析和对抗式神经网络模型的相关工作。第 2 节详细描述本文提出的对抗式神经网络模型以及对抗学习算法。第 3 节介绍 EMOBANK 语料库以及实验设置,并对实验结果进行详细分析。最后,第 4

节给出本文的结论并展望未来工作。

## 1 相关工作

本节首先介绍情绪分类与情绪回归的相关研究,然后给出对抗式神经网络模型的相关研究工作。

### 1.1 情绪分类研究

情绪分类研究的主流为基于语料库的方法.随着互联网的发展,许多研究者通过各类在线社交平台用户发布的文本构建情绪语料库.Mishne 等人<sup>[1]</sup>从在线博客平台 LiveJournal 的 815 494 篇博客文章中构建情绪分类语料库.Pak 等人<sup>[26]</sup>从 Twitter 用户发布的推文(tweets)中构建情绪分类语料库.Yao 等人<sup>[2]</sup>和 Huang 等人<sup>[8]</sup>分别从新浪微博与腾讯微博中构建中文情绪分类语料库.除情绪分类语料库的建设之外,情绪词典(emotion lexicon)的构建也是重要的语料库构建工作.Xu 等人<sup>[27]</sup>利用同义词词典、语义词典等语言资源,通过基于图规则的方法构建包含 5 类情绪的情绪词典.Yang 等人<sup>[28]</sup>提出了一种情绪意识隐狄利克雷分布模型(emotion-aware LDA model)来构建特定领域的情绪词典.

根据输入文本的粒度,情绪分类任务可以分为文档级(document-level)情绪分类和句子级(sentence-level)情绪分类.基于传统方法的文档级情绪分类包括以下代表性工作:Mishne 等人<sup>[1]</sup>利用 LiveJournal 的博客文章训练 SVM 分类器来对博客文档情绪进行分类.Yang 等人<sup>[5]</sup>借助情感词典等外部语言资源训练 SVM 分类器与 CRF 序列标注器在博客文章上进行情绪分类任务.Lin 等人<sup>[29]</sup>在雅虎新闻(Yahoo! News)上进行读者情绪分类的研究.新闻的读者情绪由新闻末尾的读者情绪投票结果确定.基于传统方法的句子级情绪分类主要依靠情绪词典确定句子包含的情绪,包括以下的代表性工作:Amam 等人<sup>[30]</sup>提出了一种基于知识库的句子级情绪识别方法.Mohammad 等人<sup>[31]</sup>对情绪词对句子情绪的影响进行了研究.他们使用了基于 Word Net 与 NRC-10 语料的情绪词典辅助训练最大熵分类器与 SVM 分类器,用以识别句子的情绪.Das 等人<sup>[32]</sup>在 Bengali 博客上识别句子级的情绪.他们首先识别句子中所包含的情绪词,之后利用情绪词典确定这些情绪词所属的情绪类别,最后综合确定句子的情绪.

近年来,基于神经网络的方法也被应用到情绪分类的任务中去.Bertero 等人<sup>[33]</sup>提出一种基于卷积神经网络(convolutional neural network,简称 CNN)的模型来对交互对话系统中的实时对话情绪进行分类.文献[33]给出的实验结果表明,使用单层卷积神经网络配合最大池化(max-pooling)的特征抽取方法便能够取得较好的分类准确率.Felbo 等人<sup>[34]</sup>提出了一种基于双向长短期记忆网络(bi-directional long short term memory,简称 Bi-LSTM)的方法在 8 个基准数据集上进行情绪分类任务研究.除了双向长短期记忆网络之外,Felbo 等人还利用推特中的情绪表情符号辅助训练模型,使之能够从文本中抽取出包含更丰富的情绪信息的特征,并在所有基准数据集上都取得了先进的性能.Abdul-Mageed 等人<sup>[3]</sup>提出了一种基于循环神经网络(recurrent neural network,简称 RNN)的情绪分类模型 EmoNet.该方法结合了循环神经网络与远程监督(distant supervision)方法对未标注数据进行自动标注,并在他们自己收集的包含 24 类情绪的语料上取得了 87.58%的准确率.

上述情绪分类研究大多采用不同的情绪分类语料库,而这些语料库往往使用不同的情绪分类体系.因此使用这些方法训练的分类模型无法直接应用到其他语料库上,而是需要在新的语料上重新训练模型.此外,由于情绪分类任务本质上是将文本特征通过模型映射到离散的标签空间中去,因此,若所使用的情绪标签种类较少,则无法进一步进行更加细粒度的情绪分析.大量增加情绪标签会大幅度提高每一类数据的稀疏性,而克服这一稀疏性需要耗费大量的人力标注成本.

### 1.2 情绪回归研究

与情绪分类任务相比,受限于回归任务本身的难度和情绪回归语料库的缺乏,情绪回归任务研究的起步较晚.Yu 等人<sup>[35]</sup>实现了一种基于带权图(weighted graph)的词语级情绪回归模型.该方法通过带权图来建模多个情绪词节点之间的关系与相似度,从而给情感词的极性-强度分数进行打分.该方法在中文和英文的情感词典上的词语级情绪回归任务中的性能均优于线性回归(linear regression)、核方法(kernel method)以及 Pagerank 算

法.Wang 等人<sup>[36]</sup>提出了一种局部卷积神经网络-长短期记忆网络(regional CNN-LSTM)的文档级情绪回归模型.该方法将一段文本划分为多个局部,并通过卷积神经网络抽取每个局部的特征.抽取得到的特征经过融合后,经由一个长短期记忆网络来预测整个文本的极性-强度分数.在多个语料上的实验结果表明,该方法优于传统的基于词典的方法以及基于单层卷积神经网络和单层长短期记忆网络的方法.Buechel 等人<sup>[37]</sup>在 SemEval 07 语料上探讨了将极性-强度-可控性分数映射为 SemEval 07 语料所使用的 Ekman 这 6 类情绪类别标签的可行性.他们首先人工地为 SemEval 07 语料标注了极性-强度-可控性分数,之后通过  $k$  近邻算法构建情绪分数和情绪标签之间的映射,并取得了较好的准确率.

在相关语料库的构建方面,现有的情绪回归语料库数量极少,并且语料库规模大多较小.Preotiuc-Pietro 等人<sup>[38]</sup>收集 Facebook 用户发布的信息,构建了规模为 2 895 条文本的英文情绪回归语料库.Yu 等人<sup>[39]</sup>发布了一个规模为 2 009 条句子的中文情绪回归语料库.该语料库的文本数据源自多个在线平台.由 Buechel 等人<sup>[4]</sup>标注的 EMOBANK 语料库是目前罕有的包含样本数超过 10 000 的大规模多领域情绪回归语料库.

与情绪分类相比,情绪回归所使用的语料库在标注上都采用心理学上公认的极性-强度模型或极性-强度-可控性模型.因此,这些方法训练的模型具有较好的普适性.此外,由于情绪回归任务本质上是将文本特征通过模型映射到连续的情绪分数空间中去,因此适合进行更细粒度的情绪分析任务.

### 1.3 对抗式神经网络研究

基于神经网络的对抗式学习方法是目前深度学习的热点之一.这类方法最早源自 Goodfellow 等人<sup>[19]</sup>提出的 GAN 模型.GAN 模型由两部分组成:生成器  $G$  和判别器  $D$ ,并使用对抗式学习算法来训练这两部分. $G$  的输入通常为的一组满足正态分布的随机噪音,输出为一张图片; $D$  的输入为一张图片,输出为判别该图片为原始数据集的真实样本还是由  $G$  生成的伪样本.对抗式学习发生在  $G$  与  $D$  之间: $G$  需要生成尽可能能够欺骗  $D$  的图片,而  $D$  需要尽可能判别出  $G$  所生成的样本为伪样本.GAN 的目标在于通过对抗式学习使得  $G$  能够生成尽可能接近真实样本分布的伪样本.当两个网络之间达到纳什均衡时模型收敛.GAN 的优点在于  $G$  的输入可以任意采样,并且整个模型在实现上可以使用任何可微模型,但同时也存在众多缺陷.在 GAN 的基础上,研究者们尝试解决该方法的不足,许多衍生模型被提了出来,并在图像处理领域取得了较好的效果.Mirza 等人<sup>[20]</sup>提出了条件式 GAN(conditional GAN,简称 CGAN)来解决 GAN 生成样本不具可控性的缺陷.CGAN 在  $G$  和  $D$  的输入上分别增加了一个条件信号作为辅助输入以实现生成数据的控制.CGAN 在 MNIST 手写数字图像数据集上用 0~9 的数字符号作为辅助的条件输入,通过实验表明这一改进能够生成指定的质量较好的手写数字图像.Radford 等人<sup>[21]</sup>探索了  $G$  与  $D$  的具体实现,通过穷举法发现,分别用 4 层卷积神经网络实现的  $G$  与  $D$  能够获得较好的生成效果,并在 MNIST 和亚洲人脸数据集(Asian face dataset)上通过实验证明了这一点.Arjovsky 等人<sup>[40]</sup>提出的 Wasserstein GAN(WGAN)则进一步解决了 GAN 训练不稳定、缺少一个具体的数值来指示训练进程的缺陷.WGAN 最主要的改进在于使用真实样本分布于生成样本分布之间的 Wasserstein 距离来取代原始 GAN 中使用的两个分布之间的简森-香农散度作为判别器  $D$  的损失函数,从而避免了当两个分布之间没有重叠或重叠很小时, $D$  的损失函数值无法度量两个分布之间的真实距离.在具体实现中,两个分布之间的 Wasserstein 距离通过一个全连接层的神经网络来近似拟合而得到.

除了图像生成之外,近年来,基于神经网络的对抗式学习算法也被应用到其他任务中去.在自然语言处理领域,该方法在文本生成的任务上取得了一定的研究进展.Zhang 等人<sup>[22]</sup>、Zhao 等人<sup>[23]</sup>均使用卷积神经网络、长短期记忆网络等模型构建 GAN 进行文本生成实验,并能够生成比较接近自然语言的文本语句.除了文本生成任务外,该方法还被应用到分类任务上.Wu 等人<sup>[24]</sup>分别使用卷积神经网络与循环神经网络构造两种 GAN 模型,在纽约时报语料与华盛顿大学语料两个数据集上进行关系抽取实验,并探讨了这两种实现的优劣.Liu 等人<sup>[25]</sup>利用对抗式学习框架构建了两个基于长短期记忆网络的分类模型进行多任务文本分类.对抗式学习发生在两个分类网络之间.在 Liu 等人收集的 16 个不同的文本分类数据集上的实验证明了这一方法在分类准确率上超过了其他单任务的深度神经网络模型.近两年来,有研究者利用对抗式学习算法的特点,将其用于跨领域的文本分类任务之中.Li 等人<sup>[41]</sup>与 Peng 等人<sup>[42]</sup>先后提出基于对抗式学习算法的神经网络模型,利用判别器对输入

文本所属领域进行判别,以此从来自源领域的输入文本中学习更接近目标领域的特征,用于训练针对目标领域测试集的文本分类器.这两种方法均在 Yelp 多领域情感分类数据集上取得了优于传统跨领域学习的分类性能.

## 2 基于对抗式神经网络的多维度情绪回归方法

本节详细描述了我们所提出的基于对抗式神经网络的多维度情绪回归方法.首先,给出多维度情绪回归任务的定义.接着,介绍本文提出的对抗式神经网络模型的框架与每一部分的具体实现,包括特征抽取器、回归器以及判别器的具体实现.最后,介绍本文在训练模型时所使用的对抗式学习算法.

### 2.1 任务定义

本文在 EMOBANK 多领域情绪回归语料上进行多维度情绪回归任务.首先,记整个语料库为  $X=\{x_1, x_2, \dots, x_n\}$ ,其中, $n$ 为语料库容量.对于语料库中第  $i$  个文本输入  $x_i$ ,本文的模型最终输出该文本的包含极性-强度-可控性情绪分数的三元组  $y_i=(V_i, A_i, D_i)$ ,其中, $x_i$  为由第  $i$  个文本中词语的词向量序列组成的词向量矩阵, $V_i$ 、 $A_i$ 、 $D_i$  分别对应文本  $x_i$  包含的极性(valence)分数、强度(arousal)分数、可控性(dominance)分数.最后,记整个语料库  $X$  对应的输出为  $Y=\{y_1, y_2, \dots, y_n\}$ .

关于词向量的训练,本文在 EMOBANK 语料库上使用 Skip-gram 算法<sup>[43]</sup>预先训练词向量查找表  $E \in \mathbb{R}^{d \times V}$ ,其中, $d$ 为词向量的维度, $V$ 为 EMOBANK 包含的单词的集合.对于第  $i$  个文本,文本中的每个单词通过查表的方式转换为词向量:

$$e_i = Ew_i \quad (1)$$

其中, $e_i$ 为第  $i$  个词对应的词向量, $w_i$ 为第  $i$  个词的单热点表示.此时可得到词向量矩阵  $x_i=[e_1, e_2, \dots, e_k]^T$ ,其中, $k$ 为每个文本的固定序列长度.对于长度小于  $k$  的文本,本文采用随机正态初始化的方法构建不足部分的词向量添加于序列首部;对于长度大于  $k$  的文本,本文从序列尾部开始截断超出的部分. $k$  作为模型的超参,其具体设置将在实验部分另作介绍.

### 2.2 对抗式神经网络模型框架

本文实现了一种对抗式神经网络模型,用于实现 3 个维度之间的两两对抗学习.以极性和强度两个维度间的对抗为例,模型的框架图如图 1 所示.

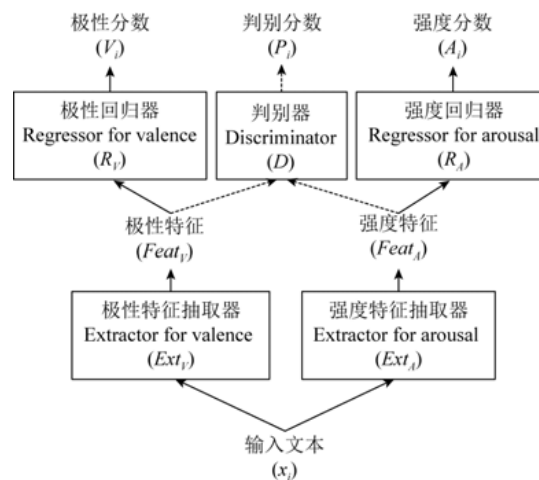


Fig.1 The framework of the proposed adversarial neural network for the adversarial learning between a pair of emotion dimensions. Taking the pair of valence and arousal as an example

图 1 用于实现一对情绪维度之间的对抗学习的神经网络模型框架图.以极性和强度两个维度为例

模型由 3 部分组成:特征抽取器( $Ext_v$ 、 $Ext_a$ )、回归器( $R_v$ 、 $R_a$ )、判别器( $D$ ).特征抽取器用于抽取文本中针

对不同维度的特征向量,抽取的特征向量被输入到相应的回归器中以对情绪维度进行打分.为了抽取具有维度特定性且泛化性更好的特征,两个维度的特征抽取器之间通过判别器进行对抗式学习.判别器用于判断抽取的特征属于何种维度.特征抽取器不仅通过常规的有监督训练来学习抽取具有维度特性的特征,还通过欺骗判别器来学习抽取泛化性更好的特征.

除图 1 所示的极性-强度对抗模型之外,本文还实现了极性-可控性、强度-可控性的对抗模型.其中包括输出可控性特征  $Feat_D$  的可控性回归器  $Ext_D$ 、输出可控性分数  $D_i$  的可控性回归器  $R_D$ .判别器  $D$  的输出  $P_i \in (-1,1)$  为输入的特征属于两个维度中某一个的判别分数.本文在此分别定义了 3 个对抗模型中判别器  $D$  输出的判别分数  $P_i$  如下:

1) 在极性-强度对抗模型中,若  $P_i$  越接近 1,则代表输入特征越可能是极性特征;反之,若  $P_i$  越接近-1,则代表输入特征越可能是强度特征.

2) 在极性-可控性对抗模型中,若  $P_i$  越接近 1,则代表输入特征越可能是极性特征;反之,若  $P_i$  越接近-1,则代表输入特征越可能是可控性特征.

3) 在强度-可控性对抗模型中,若  $P_i$  越接近 1,则代表输入特征越可能是强度特征;反之,若  $P_i$  越接近-1,则代表输入特征越可能是可控性特征.

为方便起见,在本节剩下的内容中,若不作特殊说明,则均以极性-强度模型为例来解释模型的详细实现.

### 2.3 基于长短期记忆网络的特征抽取器

自然语言处理领域中的许多研究表明,长短期记忆网络(long short term memory,简称 LSTM)能够从文本中抽取富含信息量的特征<sup>[3,35,37]</sup>.Hochreiter 等人<sup>[44]</sup>提出的 LSTM 是神经网络的一个变种模型.由于传统的循环神经网络在反向传播梯度时存在梯度消失(gradient vanishment),因此,当输入序列较长时,传统循环神经网络会丢失一部分靠前的序列中的信息.LSTM 在循环神经网络的基础上引入 LSTM 单元(LSTM cell)来解决梯度消失的问题.LSTM 单元由 4 部分组成:输入门(input gate) $i$ 、输出门(output gate) $o$ 、遗忘门(forget gate) $f$  以及记忆单元(memory cell) $c$ .给定词向量矩阵  $x_t=[e_1, e_2, \dots, e_k]^T$ ,在  $t$  时刻 LSTM 单元的计算如下:

$$i_t = \sigma(W_i e_t + U_i h_{t-1} + V_i c_{t-1}) \quad (2)$$

$$o_t = \sigma(W_o e_t + U_o h_{t-1} + V_o c_{t-1}) \quad (3)$$

$$f_t = \sigma(W_f e_t + U_f h_{t-1} + V_f c_{t-1}) \quad (4)$$

$$\tilde{c}_t = \tanh(W_c e_t + U_c h_{t-1}) \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

其中, $e_t$  为  $t$  时刻 LSTM 单元的输入,亦即输入矩阵  $x$  中第  $t$  个行向量, $h_t$  为  $t$  时刻 LSTM 单元的输出, $\sigma$  表示 sigmoid 激活函数, $\odot$  表示点乘运算符, $W$ 、 $U$ 、 $V$  均表示 LSTM 单元的参数矩阵.输入门控制每个记忆单元的更新信息.输出门控制每个记忆单元的输出信息.遗忘门控制每个记忆单元在时间  $t$  时刻丢弃的信息. $\tilde{c}_t$  表示  $t$  时刻记忆单元的候选值. $c_t$  根据  $t-1$  时刻的记忆单元状态  $c_{t-1}$  和  $t$  时刻的候选值  $\tilde{c}_t$  计算得到.

作为模型的原始输入,词向量矩阵  $x_t$  输入 LSTM 之后,计算得到输出序列  $H=[h_1, h_2, \dots, h_k]^T$ .由于最后一个时刻  $k$  的计算实际上取决于时刻  $1 \sim k-1$  的输出,因此,在过去的研究中大多选用  $h_k$  作为 LSTM 所抽取的序列中的特征向量.为了从文本中获取更加丰富的信息,本文另对输出序列  $H$  进行平均池化(average pooling)操作.

$$\bar{h} = \frac{1}{k} \sum_{t=1}^k h_t \quad (8)$$

平均池化后得到的输出向量与  $h_k$  进行拼接操作后通过  $\tanh$  激活函数得到特征抽取器最终输出的特征向量  $Feat$  为

$$Feat = \tanh(h_k \oplus \bar{h}) \quad (9)$$

图 2 给出了特征抽取器的神经网络结构.在极性-强度模型中,特征抽取器抽取词向量矩阵  $x_t$  中的极性特征与强度特征的计算流程定义如下:

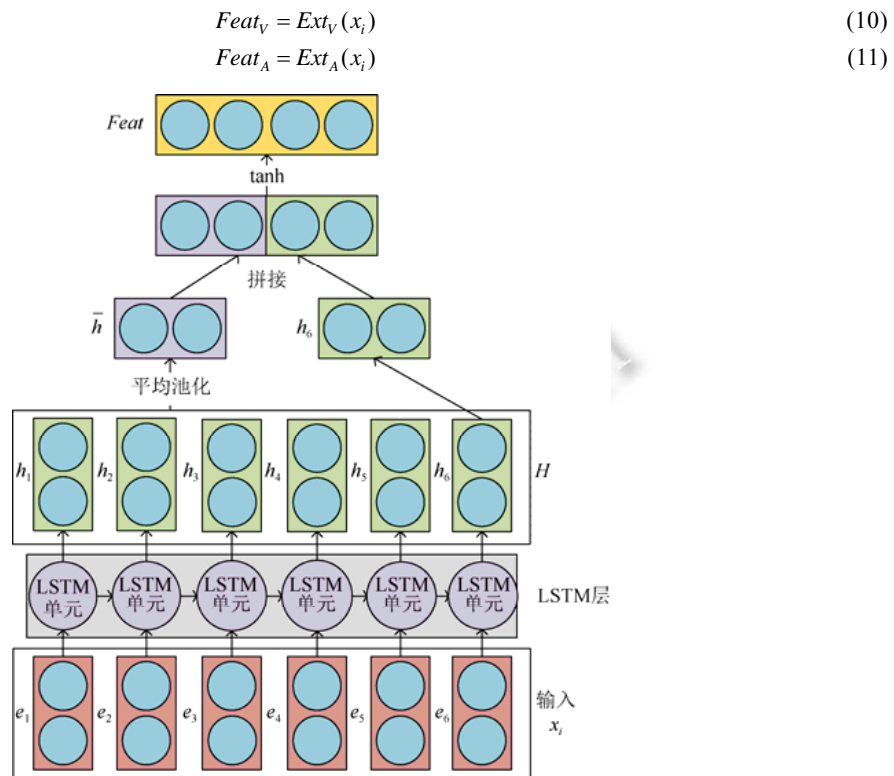


Fig.2 The architecture of the LSTM-based neural network for extracting feature vector from the input text

图 2 用于抽取输入文本的特征向量的基于 LSTM 神经网络的抽取器结构

## 2.4 回归器

回归器接收特征抽取器所抽取的输入文本特征作为输入,用于对输入文本某个情绪维度的分数进行打分.本文所提出的对抗式模型并不限制回归器部分的具体实现,因此,只要模型每个部分的梯度能够在网络中传递,回归器既可以是基于传统方法的回归器,也可以是任何可微神经网络模型构成的回归器.本文为了突出对抗式学习在训练特征抽取器上的优越性,仅使用简单的单层全连接神经网络层作为回归器的实际实现方式.

$$h = W \cdot Feat + b \quad (12)$$

$$S = \text{relu}(h) \quad (13)$$

其中, $W$ 表示全连接层的参数, $b$ 为偏置项, $h$ 为全连接层的隐层状态, $\text{relu}$ 表示 Relu 激活函数.最终输出为某个情绪维度的分数  $S$ .在极性-强度模型中,回归器的计算流程定义如下:

$$S_V = R_V(Feat_V) \quad (14)$$

$$S_A = R_A(Feat_A) \quad (15)$$

## 2.5 判别器

判别器接收特征抽取器输出的文本中的特征作为输入,以判别抽取的特征是属于哪个情绪维度的特征.判别器在对抗式学习算法中起着重要的作用,模型不同部分之间的对抗均通过判别器来进行.在判别器的实现上,本文参照了 Arjovsky 等人<sup>[40]</sup>在 WGAN 上的工作,使用两个特征分布之间的 Wasserstein 距离来指示判别器的训练程度.相比于原始 GAN 中使用 JS 散度作为判别器的损失函数,Wasserstein 距离在两个分布没有重叠或重叠很小的时候也能够度量两个分布之间的距离,从而为判别器的训练提供更加平滑的度量来衡量模型的训练水平.在实际实现中,本文使用单层全连接层来近似拟合 Wasserstein 距离.

$$h = W \cdot Feat + b \quad (16)$$

$$P = \tanh(h) \quad (17)$$

其中,  $W$  表示全连接层的参数,  $b$  为偏置项,  $h$  为全连接层的隐层状态,  $\tanh$  表示双曲正切激活函数. 最终输出为判别结果  $P$ . 在极性-强度模型中, 判别器的计算流程定义如下:

$$P = D(\text{Feat}_V) \quad (18)$$

$$P = D(\text{Feat}_A) \quad (19)$$

## 2.6 对抗式学习算法与模型训练

本文提出一种对抗式学习算法来训练对抗式神经网络. 该算法首先通过最小化以下的回归损失函数来训练特征抽取器  $\text{Ext}_V$  和  $\text{Ext}_A$ 、回归器  $R_V$  和  $R_A$ . 在具体实现上, 本文采用均方误差作为回归的损失函数.

$$\min \frac{1}{N} \sum_{i=1}^N (\tilde{S}_i - S_i)^2 = \min \frac{1}{N} \sum_{i=1}^N (R(\text{Ext}(x_i)) - S_i)^2 \quad (20)$$

其中,  $N$  为样本总数,  $\tilde{S}_i$  和  $S_i$  分别为第  $i$  个样本情绪分数的估计值和真实值. 在训练完特征抽取器和回归器之后, 该算法通过最大化以下判别误差(即 Wasserstein 距离)来训练判别器  $D$ .

$$\max \frac{1}{N} \sum_{i=1}^N [D(\text{Feat}_{V_i}) - D(\text{Feat}_{A_i})] \quad (21)$$

其中,  $\text{Feat}_{V_i}$  和  $\text{Feat}_{A_i}$  分别代表第  $i$  个样本的极性特征和强度特征. 最后, 对抗式学习算法通过最小化以下判别误差达到欺骗判别器的效果, 从而再次训练特征抽取器  $\text{Ext}_V$  和  $\text{Ext}_A$ .

$$\min \frac{1}{N} \sum_{i=1}^N [D(\text{Feat}_{V_i}) - D(\text{Feat}_{A_i})] \quad (22)$$

算法 1 给出了详细的在一轮学习中使用对抗式学习训练整个网络的算法. 模型在若干轮学习后收敛. 其中, 该算法在训练判别器时将判别器的参数的绝对值进行了裁剪. 根据 Arjovsky 等人<sup>[40]</sup>在 WGAN 上的工作, 进行判别器的参数裁剪是为了使判别器这一可微网络满足 Lipschitz 连续性(Lipschitz continuity)条件, 从而能够使用该网络的输出结果来近似拟合 Wasserstein 距离.  $t$  作为模型的可调超参, 它的取值将在本文实验部分给出.

**算法 1.** 对抗式学习算法.

- (1) 从训练集中顺序选取  $N$  条训练数据  $X_N = \{x_1, x_2, \dots, x_N\}$ .
- (2) 利用  $X_N$ , 根据训练样本所标注的两个维度的情绪分数分别训练特征抽取器  $\text{Ext}_V$  和  $\text{Ext}_A$ 、回归器  $R_V$  和  $R_A$ . 更新  $\text{Ext}_V$ 、 $\text{Ext}_A$ 、 $R_V$ 、 $R_A$  的参数.
- (3) 利用  $X_N$ , 训练判别器  $D$ . 更新  $D$  的参数.  $\text{Ext}_V$ 、 $\text{Ext}_A$  的参数保持不变.
- (4) 将更新后  $D$  的参数的绝对值裁剪到一个不大于  $t$  的值.
- (5) 利用  $X_N$ , 通过欺骗  $D$  再次训练  $\text{Ext}_V$  和  $\text{Ext}_A$ . 更新  $\text{Ext}_V$  和  $\text{Ext}_A$  的参数.  $D$  的参数保持不变.
- (6) 重复(1)~(5), 直至整个训练集的样本被取完.

在更新网络参数中, 对于模型不同的部分, 本文采用不同的优化算法. 对于特征抽取器和回归器, 本文使用 Adam 算法<sup>[45]</sup>作为梯度优化算法. 判别器则根据 RMSProp 算法<sup>[46]</sup>来更新它的权重. 神经网络模型中每一层的参数都采用位于区间  $[-\sqrt{6/(r+c)}, \sqrt{6/(r+c)}]$  上的正态取样进行初始化, 其中,  $r$  与  $c$  分别表示参数矩阵的行数与列数<sup>[47]</sup>.

## 3 实验

本节首先介绍所使用的情绪回归语料库 EMOBANK. 之后给出实验的设置, 包括衡量情绪回归性能的目标、模型中的超参设置. 然后介绍实验中所使用的基准系统, 以与本文提出的方法进行对比实验. 本文使用的外部工具也在本节中相应部分给出. 最后, 给出详细的实验结果, 包括具体的回归性能与结果分析.

### 3.1 EMOBANK语料库

本文使用 EMOBANK 语料库<sup>[4]</sup>来验证我们所提出的基于对抗式神经网络的多维度情绪回归方法. 该语料



库的数据源自两个公开语料库:SemEval07:task 14(<http://nlp.cs.swarthmore.edu/semeval/tasks/task14/data.shtml>)与 MASC(<http://www.anc.org/data/masc/corpus/>)语料库,人工标注了共 6 个领域的英文文本的 10 325 条读者情绪和 10 279 条作者情绪,由多个标注者分别对文本的 3 个情绪维度:极性、强度、可控性进行读者情绪和作者情绪打分,分数区间为[1.0,5.0].EMOBANK 中的读者情绪和作者情绪的文本数量并不相同,这说明语料库在标注时有少量文本未作读者情绪或作者情绪的标注.表 1 给出了 EMOBANK 包含的文本数量在各领域上的分布.由于本文不涉及跨领域情绪回归的研究,而不同领域文本的相同情绪维度的特征分布存在一定的不同,因此本文仅在样本数最多的两个领域:新闻领域与小说领域上分别进行实验.

**Table 1** Genre distribution of EMOBANK corpus

**表 1** EMOBANK 语料各领域文本数量的分布

领域	读者情绪	作者情绪
新闻	2 560	2 540
博客	1 364	1 349
散文	1 182	1 238
小说	2 824	2 819
信件	1 445	1 383
旅游指南	950	950
合计	<b>10 325</b>	<b>10 279</b>

表 2 给出了 EMOBANK 语料库的主要统计数据.从表中可以看出,无论是读者情绪还是作者情绪,3 个情绪维度的标注分数平均值都在 3.00 上下.而情绪分数的标准差较低,样本的情绪分数大多落于区间[2.3,3.7]上.

**Table 2** Statistics of EMOBANK corpus

**表 2** EMOBANK 语料库的统计数据

		读者情绪	作者情绪
文本长度统计	平均文本长度	14.98	15.01
	文本长度标准差	11.33	11.40
	平均极性分数	2.97	2.98
情绪分数统计	极性分数标准差	0.43	0.34
	平均强度分数	3.05	3.03
	强度分数标准差	0.33	0.32
	平均可控性分数	3.05	3.08
	可控性分数标准差	0.29	0.26
显著性检验 (配对双样本 <i>t</i> 检验)	极性 vs. 强度	$p=3.43e-54$	$p=1.47e-27$
	极性 vs. 可控性	$p=0.24$	$p=2.55e-159$
	强度 vs. 可控性	$p=8.13e-81$	$p=3.04e-52$

EMOBANK 中文本平均长度约为 15 个单词,而文本长度的标准差较大.由于本文提出的神经网络模型的输入需要保证每个词向量矩阵的行列相等,因此会预先将文本长度调整到等长.为了既不丢失绝大多数文本中的信息,也不在较短文本中加入过多的空白填充,本文在实验中将文本长度  $k$  统一设置为 37(均值+两倍标准差).

除了文本长度与情绪分数的计量之外,本文还使用了配对双样本 *t* 检验来考察文本两两情绪维度分数之间是否存在显著差异.从表 2 中可以看出,除了极性与可控性两个维度的读者情绪之外,剩余两两组合之间的差异性均相当明显.不同维度情绪分数间的明显差异说明了在 EMOBANK 上的多维度的情绪回归任务本身具有意义.

### 3.2 实验设置

本文在 EMOBANK 的新闻与小说领域上分别进行读者情绪与作者情绪回归的 5 倍交叉验证,并使用皮尔森相关系数(Pearson's correlation coefficient,用  $r$  表示)与均方误差(mean square error,简称 MSE)作为评估指标.

$$r = \frac{\sum_{i=1}^N (\tilde{S}_i - \bar{\tilde{S}})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^N (\tilde{S}_i - \bar{\tilde{S}})^2} \sqrt{\sum_{i=1}^N (S_i - \bar{S})^2}} \quad (23)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (\tilde{S}_i - S_i)^2 \quad (24)$$

若系统在测试集上的  $r$  越高, MSE 越低, 则说明系统性能越好. 本文在评估各方法性能时优先考虑皮尔森相关系数  $r$ , 而 MSE 作为辅助参考指标. 此外, 本文使用配对双样本  $t$  检验来检验本文方法与基准系统性能之间的差异是否显著. 实验中各超参的设置见表 3 中给出. 本文使用 gensim 工具包(<https://radimrehurek.com/gensim/>)的 word2vec 训练词向量查找表, 使用 skipgram 算法在新闻和小说领域的语料上分别进行无监督训练.

**Table 3** Setting of hyper parameters during experimentation

表 3 实验中超参的设置

	超参名	超参值
词向量训练	词向量维度	300
	迭代次数	10
	窗口大小	3
神经网络训练	LSTM 层输出维度	300
	特征抽取器学习速率	2e-4
	回归器学习速率	4e-4
	判别器学习速率	2e-4
	判别器参数裁剪值 $t$	0.1
	输入文本长度 $k$	37

### 3.3 基准系统

为充分证明我们提出的方法在 EMOBANK 上进行多维度情绪回归任务的优越性, 本文进行了以下基准系统与所提方法的比较.

1) SVR(BOW): 该基准系统使用了 Smola 等人<sup>[48]</sup>提出的基于支持向量机的回归器(support vector regressor, 简称 SVR). SVR 被广泛地应用到自然语言处理的回归任务中, 并被认为是该领域的先进回归系统之一<sup>[49]</sup>. 该系统使用文本的词袋模型(bag-of-words)作为输入特征.

2) SVR(TF-IDF): 该系统与基准系统 1) 相同, 使用 SVR 作为回归器. 与 1) 的不同之处在于, 该系统以文本中单词的 TF-IDF 度量作为输入特征.

3) SVR(W2V): 该系统与基准系统 1) 相同, 使用 SVR 作为回归器. 与 1) 的不同之处在于, 该系统以文本中单词的词向量之和作为输入特征. 该系统所使用的词向量查找表与本文提出的方法相同.

4) ANN: 该系统由 Bitvai 等人<sup>[50]</sup>提出. 其基本结构为使用多层并联的卷积神经网络抽取文本中的多个  $n$  元文法( $n$ -gram)特征. 该方法被应用于对电影评论的情感极性回归任务中, 亦被认为是自然语言处理领域中的先进回归系统之一. 本文所实现的该系统采用 3 层并联卷积神经网络, 分别抽取文本中的一元文法(uni-gram)、二元文法(bi-gram)、三元文法(tri-gram)特征. 抽取的 3 种特征经过拼接后输入到一层全连接层来进行回归任务. 该系统的输入特征与本文提出的方法相同, 为文本的词向量矩阵.

5) LSTM: 该系统为基于长短期记忆网络的回归器. 与本文提出的方法相比, 该系统仅使用图 2 所描述的模型抽取文本特征, 再以抽取的文本特征作为输入, 使用第 2.4 节所描述的回归器进行文本回归任务. 该系统的输入特征与本文提出的方法相同, 为文本的词向量矩阵.

6) C-LSTM: 该系统由 Zhou 等人<sup>[51]</sup>提出, 为文本分类任务上的先进系统. 本文通过修改模型最后一层的激活函数, 将该方法应用到文本回归的任务中. 该方法组合了卷积神经网络以及长短期记忆网络, 首先利用卷积神经网络将句子转换为短语级特征, 再将短语级特征输入到长短期记忆网络中进行分类任务. 该系统的输入特征与本文提出的方法相同, 为文本的词向量矩阵.

此外, 本文提出的对抗式神经网络简称为 AdvNN.

### 3.4 实验性能对比与分析

表 4 给出了 AdvNN 与其他基准系统在新闻领域上的情绪回归性能. 由于本文实验采用 5 倍交叉验证, 评估结果为每次验证结果的平均值. 此外, 表 5 进一步给出了 3 种情绪维度两两组合: 极性-强度、极性-可控性以及强

度-可控性 3 个模型的详细性能.以极性-强度模型为例,表中维度 1 表示前一维度——极性维度回归性能,维度 2 表示后一维度——强度维度回归性能.表 4 中 AdvNN 的每个维度的性能取自表 5 中各组合中皮尔森相关系数  $r$  最高的一项,例如读者情绪的极性维度在极性-可控性模型中的回归结果的  $r$  值较极性-强度模型中的  $r$  值更高,因此,表 4 中记录该最优  $r$  值以及相应的 MSE 值.

**Table 4** The performances of various approach on the News domain  
**表 4** 各种方法在新闻领域上的情绪回归性能

基准系统		极性		强度		可控性	
		$r$	MSE	$r$	MSE	$r$	MSE
读者情绪	SVR(BOW)	0.309	0.206	0.130	0.098	0.116	0.086
	SVR(TF-IDF)	0.307	0.212	0.139	0.094	0.155	0.094
	SVR(W2V)	0.272	0.198	0.144	0.094	0.139	0.090
	ANN	0.288	0.219	0.150	<b>0.088</b>	0.136	0.091
	LSTM	0.324	0.205	0.127	0.098	0.147	0.090
	C-LSTM	0.337	0.190	0.143	0.094	0.168	0.087
	AdvNN( $r$ 最优)	<b>0.353</b>	<b>0.182</b>	<b>0.168</b>	0.091	<b>0.179</b>	<b>0.080</b>
作者情绪	SVR(BOW)	0.251	0.114	0.092	0.093	0.129	0.059
	SVR(TF-IDF)	0.263	0.122	0.065	0.092	0.112	0.058
	SVR(W2V)	0.242	0.121	0.081	0.095	0.143	0.055
	ANN	0.217	<b>0.111</b>	0.060	<b>0.084</b>	0.127	0.057
	LSTM	0.225	<b>0.111</b>	0.074	0.087	0.121	0.061
	C-LSTM	0.242	0.129	0.091	0.091	0.148	0.068
	AdvNN( $r$ 最优)	<b>0.290</b>	0.120	<b>0.093</b>	0.085	<b>0.160</b>	<b>0.052</b>

从表 4 可以看出:

1) 总体而言,在新闻领域上,所有方法取得的  $r$  值都较低.最高的  $r$  值为 AdvNN 在读者极性情绪上取得的 0.353.作为参考,EMOBANK 全领域上标注者两两间标注结果的平均  $r$  值约为 0.6<sup>[4]</sup>.由此可见,EMOBANK 上的多维度情绪回归任务是一个较难的任务,不容易取得非常高的性能.注意到该领域上的作者强度情绪最高  $r$  值未超过 0.1.造成这一结果的可能原因在于新闻领域文本的作者撰写相关文本时倾向于不代入较强的情绪.例如,例句 3 的作者强度情绪为 3.25,为中等强度,而读者强度情绪为 4.00,为较高强度.例句 3 作为一条典型的报道负面消息的新闻文本,能够激发起读者较强的负面情绪.而作者在撰写这条文本时为了保证新闻的客观性,相比读者不会代入强烈的情绪(尽管例句 3 的作者情绪极性也为负面).同样,例句 4 作为一条报道正面消息的新闻文本,其读者强度情绪为较高的 3.80,而作者强度情绪为中等的 3.40.例句 5 说明在读者强度情绪较弱情况(2.00)下,作者情绪依然趋于中等(2.80).可见,在新闻领域上,作者情绪倾向于为一个适中的强度.

2) 与传统方法(SVR)相比,一般的基于神经网络的回归模型表现并无明显优势.其中,ANN 在所有基准系统中的总体性能最差,各情绪维度的平均  $r$  值为 0.163,低于其他基准系统.相比之下,SVR 方法的平均  $r$  值在 0.171 上下,优于 ANN,相比 LSTM 取得的平均  $r$  值 0.174 亦无太大劣势.而 C-LSTM 取得了基准系统中最高的平均  $r$  值 0.184,说明了文本分类任务中的先进系统在回归任务上亦能取得较好的效果.

3) 在新闻领域上,无论是读者情绪还是作者情绪,AdvNN 的最高  $r$  值均超过了其他所有基准系统的  $r$  值.其中,在作者极性维度上的提升最多,较该维度上最强的基准系统 SVR(TF-IDF)提升了 0.027. AdvNN 在作者强度维度回归上的提升最小,较该维度上最强的基准系统 SVR(BOW)仅有 0.001 的提升.此外,AdvNN 在新闻领域的平均  $r$  值达到 0.207,亦超过了平均  $r$  值最高的基准系统 C-LSTM.从这一结果来看,同样是在 LSTM 的基础上实现的神经网络模型,使用了对抗性学习的 AdvNN 相比融合了卷积神经网络的 C-LSTM 在本文的任务上能够进一步学习文本中的情绪维度特征,从而提高系统性能.

4) 除  $r$  值之外,AdvNN 在 MSE 评估指标下的表现也较为优秀.除读者强度、作者极性和作者强度之外,AdvNN 均取得了最低的 MSE 值.尽管在作者强度维度上的  $r$  值提升较低,AdvNN 与 SVR(BOW)相比 MSE 下降了 0.008,但总体而言,在该维度上的性能却存在一定的提升. AdvNN 取得了平均 0.102 的 MSE 值,亦低于所有基准系统的平均 MSE 值.

例句 3: *Scam lures victims with free puppy offer.* (骗局使用提供免费的小狗的方式来引诱受害者.)

例句 4: Kidnapped AP photographer freed in Gaza. (被绑架的美联社摄影师在加扎被释放)

例句 5: I feel bad for Jorge he's a good friend of mine, Bengie said. (本吉说到: 我为霍赫感到难过, 他是我的好朋友)

**Table 5** Detailed performances of AdvNNs for three combinations of emotion dimensions on the News domain

**表 5** 新闻领域上 3 种情绪维度组合的对抗式神经网络详细性能

	维度组合相应模型	维度 1		维度 2	
		$r$	MSE	$r$	MSE
读者情绪	极性-强度模型	0.339	0.190	<b>0.168</b>	0.091
	极性-可控性模型	<b>0.353</b>	0.182	0.174	0.083
	强度-可控性模型	0.162	0.097	<b>0.179</b>	0.080
作者情绪	极性-强度模型	<b>0.290</b>	0.120	<b>0.093</b>	0.085
	极性-可控性模型	0.287	0.118	<b>0.160</b>	0.052
	强度-可控性模型	0.090	0.085	0.149	0.061

进一步分析表 5, 可以发现:

1) 除了强度-可控性模型中作者强度情绪的  $r$  值低于两个基准系统之外, 各维度组合整体结果相比基准系统均有一定的提升. 其中, AdvNN 在读者强度情绪和作者极性情绪上的非最优  $r$  值亦超过最强基准系统至少 0.019.

2) 除作者强度情绪外, 其余情绪维度在不同组合下的平均  $r$  值亦超过相应维度的最强基准系统结果. 尽管在大多数情绪维度上 AdvNN 的所有维度组合性能超过了最强基准系统, 然而由于某些维度组合的  $r$  值提升幅度较小, 如极性-强度模型的读者情绪相比该维度最强基准系统 C-LSTM 仅有 0.002 的  $r$  值提升. 因此, 尽管 AdvNN 在新闻领域上的各组合的最优性能均超过了最强基准系统, 但在个别情况下的提升并不稳定.

3) AdvNN 在作者强度情绪这一维度上的较低性能也进一步证明, 由于新闻文本本身的特性, 新闻作者在撰写相关文本时为了维持新闻的相对客观性, 并不会产生明显的情绪强度上的波动, 因而难以在该维度上取得较好的回归结果.

表 6 给出了 AdvNN 与其他基准系统在小说领域上的情绪回归性能, 具体结果亦为 5 倍交叉实验的平均值. 表 7 进一步给出了 3 种情绪维度两两组合的详细性能. 从表 6 中可以看出:

1) 与新闻领域相比, 在小说领域上, 所有方法取得的  $r$  值亦较低. 而与新闻领域不同的是, 小说领域的结果之中无很低 (小于 0.1) 的  $r$  值. 造成这一结果的原因可能在于小说作者在撰写文本时会代入一定程度的个人情绪. 例如, 例句 6 标注了负面的读者情绪和作者情绪, 而读者强度情绪和作者强度情绪分别为 4.20 与 4.20, 均为较强的情绪. 而例句 7 的读者强度情绪和作者强度情绪分别为 2.67 与 2.20, 均为较弱的情绪. 由此可见, 小说领域上的作者情绪与新闻领域上的不同, 是和读者情绪一样会产生较为明显的强度波动. 因此, 实验中所有系统相对来说在小说领域强度维度上的  $r$  值均高出新闻领域很多. 此外, 整体来看, 所有系统取得的  $r$  值与新闻领域上一样并不高, 亦证明了 EMOBANK 语料上的多维度情绪回归任务较难.

2) 不同于新闻领域, ANN 在小说领域上的性能较好, 其平均  $r$  值达到了 0.184, 在基准系统中仅次于 C-LSTM 的 0.190. LSTM 在小说领域上的表现明显差于其他基准系统, 平均  $r$  值仅为 0.160, 尤其是在读者极性维度和读者强度维度上与其他基准系统相比有较大差距. 而 SVR 在小说领域的性能依然稳定, 3 种特征输入得到的结果平均  $r$  值均在 0.170 以上, 最高为 SVR(TF-IDF) 取得的 0.182, 进一步证明了传统的 SVR 在回归任务中是不亚于一般的神经网络模型的先进系统.

3) AdvNN 在小说领域上亦取得了所有情绪维度下最高的  $r$  值, 并且在读者情绪与作者情绪的所有维度下较  $r$  值最高的基准系统有至少 0.019 的明显提升. 此外, AdvNN 在小说领域的平均  $r$  值达到了 0.219, 亦超过了平均  $r$  值最高的基准系统 C-LSTM. 在小说领域上的结果进一步证明, AdvNN 相比 C-LSTM 在本文的任务上能够进一步提升基于 LSTM 的神经网络回归模型的性能.

4) 小说领域上, AdvNN 在 MSE 评估指标下的表现也较为良好, 并在作者极性维度和作者强度维度中取得了最低的 MSE 值. 除了在读者可控性维度中的 MSE 值高出 MSE 值最低的基准系统 C-LSTM 较多之外, 在其他

未取得最低 MSE 值的情绪维度上相比该维度的最低值差距亦较小.然而与新闻领域不同的是,AdvNN 在所有情绪维度下取得了 0.124 的平均 MSE 值,与 SVR(TF-IDF)相同,略高于 C-LSTM 取得的 0.122,总体而言并未取得最佳的 MSE 值.

**Table 6** The performances of various approach on the Fictions domain

表 6 各种方法在小说领域上的情绪回归性能

基准系统		极性		强度		可控性	
		$r$	MSE	$r$	MSE	$r$	MSE
读者情绪	SVR(BOW)	0.220	0.200	0.191	0.167	0.144	0.112
	SVR(TF-IDF)	0.200	<b>0.176</b>	0.196	0.122	0.144	0.105
	SVR(W2V)	0.217	0.187	0.162	0.124	0.166	0.110
	ANN	0.228	0.181	0.201	<b>0.120</b>	0.157	0.102
	LSTM	0.185	0.188	0.132	0.143	0.152	0.102
	C-LSTM	0.229	0.179	0.202	<b>0.120</b>	0.162	<b>0.098</b>
	AdvNN( $r$ 最优)	<b>0.253</b>	0.181	<b>0.225</b>	0.122	<b>0.193</b>	0.110
作者情绪	SVR(BOW)	0.203	0.136	0.154	0.119	0.157	0.092
	SVR(TF-IDF)	0.209	0.130	0.183	0.119	0.161	0.091
	SVR(W2V)	0.199	0.135	0.132	0.120	0.155	0.087
	ANN	0.187	0.146	0.170	0.115	0.164	<b>0.085</b>
	LSTM	0.182	0.141	0.157	0.117	0.147	0.092
	C-LSTM	0.200	0.128	0.172	0.120	0.173	0.087
	AdvNN( $r$ 最优)	<b>0.243</b>	<b>0.125</b>	<b>0.202</b>	<b>0.114</b>	<b>0.199</b>	0.089

例句 6: *She screamed I haven't socialized with Terra's elite for most of my life.* (她尖叫到:我人生中大部分时间都没能与地球的精英社交.)

例句 7: *The only other illumination came from a lurid moonlight filtered through thin branches and clouds casting its bone-pale glow onto the pine floorboards.* (唯一的其他照明来自一缕透过稀薄的树枝和云彩的月光,将其苍白的的光芒投射到松木地板上.)

进一步分析表 7,可以发现:

1) 在小说领域上,AdvNN 在所有维度上的非最优  $r$  值依然高于最强基准系统,并未出现新闻领域上 AdvNN 在个别维度的非最优  $r$  值低于基准系统的情况.其中,读者极性情绪、读者可控性情绪以及作者极性情绪上 AdvNN 的非最优  $r$  值相比最强基准系统都有至少 0.018 的提升.

2) 各情绪维度在不同模型中的平均  $r$  值亦均超过相应维度的最强基准系统结果,其中,AdvNN 在作者极性情绪上取得的平均  $r$  值达到 0.239,相比该维度上的最强基准系统 SVR(TF-IDF)的提升达到非常显著的 0.030.

3) 通过上述结果可以看出,小说领域的文本由于领域的特性,其情绪分析的难度相对新闻领域较低.两个领域上的结果对比说明了读者在阅读一段包含情绪的文本时并不受领域差异性的影响,对于情绪的表达会有较明显的波动,使得情绪分析系统能够较好地从中挖掘相关的特征来完成情绪回归的任务.而作者情绪则会较明显地受领域差异性的影响,尤其是作者强度情绪.小说领域的作者情绪强度不同于新闻领域,作者不需要保持情绪上的客观性,因此小说领域的作者强度情绪和读者强度情绪一样,能够比较容易被情绪回归系统识别.

**Table 7** Detailed performances of AdvNNs for three combinations of emotion dimensions on the Fictions domain

表 7 小说领域上 3 种情绪维度组合的对抗式神经网络详细性能

	维度组合相应模型	维度 1		维度 2	
		$r$	MSE	$r$	MSE
读者情绪	极性-强度模型	<b>0.253</b>	0.181	0.211	0.128
	极性-可控性模型	0.249	0.180	<b>0.193</b>	0.110
	强度-可控性模型	<b>0.225</b>	0.122	0.184	0.119
作者情绪	极性-强度模型	0.235	0.136	<b>0.202</b>	0.114
	极性-可控性模型	<b>0.243</b>	0.125	<b>0.199</b>	0.089
	强度-可控性模型	0.189	0.119	0.182	0.094

### 3.5 显著性测试

除了上述的情绪回归实验结果之外,我们进一步给出了本文所提方法的输出结果与  $r$  值最优的基准系统输出结果之间的配对双样本  $t$  检验结果.表 8 与表 9 分别给出了新闻与小说领域中各情绪维度上双样本  $t$  检验的  $p$  值.与实验结果一样,表 8 与表 9 中的  $p$  值亦为 5 倍交叉验证每一组测试集上输出结果之间显著性测试的平均值.

**Table 8** Results of dependent  $t$ -test for paired samples between AdvNN and strongest baselines for each emotion dimension in the News domain  
表 8 新闻领域上 AdvNN 与各情绪维度下最强基准系统之间的配对双样本  $t$  检验结果

		$p$
读者情绪	极性(vs. C-LSTM)	2.65e-11
	强度(vs. ANN)	6.03e-11
	可控性(vs. C-LSTM)	3.15e-7
作者情绪	极性(vs. SVR(TF-IDF))	1.58e-23
	强度(vs. SVR(BOW))	0.082
	可控性(vs. C-LSTM)	9.10e-9

**Table 9** Results of dependent  $t$ -test for paired samples between AdvNN and strongest baselines for each emotion dimension in the Fictions domain  
表 9 小说领域上 AdvNN 与各情绪维度下最强基准系统之间的配对双样本  $t$  检验结果

		$p$
读者情绪	极性(vs. C-LSTM)	1.12e-28
	强度(vs. C-LSTM)	2.20e-30
	可控性(vs. SVR(W2V))	1.15e-28
作者情绪	极性(vs. SVR(TF-IDF))	6.33e-36
	强度(vs. SVR(TF-IDF))	6.19e-14
	可控性(vs. C-LSTM)	4.11e-21

从表 8 所示结果可以看出,在新闻领域中,除了作者强度维度之外,本文提出的 AdvNN 相比各情绪维度上最强的基准系统有显著的性能提升( $p < 0.05$ ).仅在作者强度维度上,AdvNN 相比最强基准系统 SVR(BOW)的提升不明显( $p > 0.05$ ).而表 9 的结果则表明,在小说领域中的所有情绪维度上,AdvNN 相比各维度上的最强基准系统均有显著的性能提升.显著性测试的结果表明,本文提出的 AdvNN 相比其他基准系统不仅在  $r$  值评估指标下有明显的数值上的提升,并且这一提升具有统计显著性,从而进一步证明了 AdvNN 在 EMOBANK 语料中多维度情绪回归任务上的优越性.

## 4 结论与未来工作

本文针对多维度情绪回归这一任务提出了一种基于对抗式神经网络的多维度情绪回归方法.该方法包括了对抗式神经网络和对抗式学习算法.对抗式神经网络包括 3 个部分:特征抽取器、回归器以及判别器.以在一对情绪维度上进行对抗式学习的情绪回归模型为例,首先,本文实现了两个基于 LSTM 神经网络模型的特征抽取器,以分别从输入的文本词向量矩阵中抽取与特定情绪维度相关的特征向量.其次,本文使用两个基于全连接神经网络的回归器,以分别根据特征抽取器输出的特征对特定情绪维度进行打分.最后,本文实现了一个基于全连接神经网络的判别器,以判别特征抽取器抽取的特征属于哪个情绪维度.该方法借助判别器,在两个情绪维度的特征抽取器之间通过对抗式学习算法进行对抗式训练,从而使两个特征抽取器能够从文本输入中抽取泛化性更好的针对不同情绪维度的特征向量,继而获得更好的回归性能.在 EMOBANK 语料的新闻和小说两个领域上的实验结果表明,无论是作者情绪回归还是读者情绪回归,本文的方法在 3 个情绪维度上的  $r$  值均超过了所有的基准系统,其中包括了文本回归领域中的先进系统 ANN.除了文本回归先进系统之外,本文还将文本分类领域的先进系统 C-LSTM 进行修改后用于本文针对的多维度情绪回归任务进行比较,而本文提出的方法依然在所有情绪维度上取得了更好的  $r$  值.除实验结果之外,AdvNN 与各情绪维度上最强的基准系统间的显著性测试也表明,AdvNN 取得的性能提升具有统计上的显著性,进一步证明了 AdvNN 在 EMOBANK 新闻与小说领域上的多维度情绪回归任务上的优越性.

然而,本文的方法目前仅为两个维度之间的对抗,并且在一些维度下取得的提升并不明显(例如新闻领域上的作者强度情绪维度).此外,一些情绪维度的对抗组合取得的性能提升也并不非常理想,甚至有个别组合在某个情绪维度下取得的  $r$  值相比部分基准系统有所下降.因此,我们未来的工作将首先集中于同时进行三维度之间的对抗.此外,我们还将改进对抗式学习框架与算法,例如引入共享参数的神经网络层来进一步学习抽取不同维度间的公共特征,以期能够在所有情绪维度对抗的组合上都取得稳定且明显的提升.

**References:**

- [1] Mishne G. Experiments with mood classification in blog posts. In: Shlomo A, Jussi K, James G. S, eds. Proc. of the ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access. New York: Association for Computing Machinery, 2005. 321–327.
- [2] Yao Y, Wang S, Xu R, Liu B, Gui L, Lu Q, Wang X. The construction of an emotion annotated corpus on microblog text. Journal of Chinese Information Processing, 2014,28(5):83–91 (in Chinese with English abstract).
- [3] Abdul-Mageed M, Ungar L. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In: Barzilay R, Kan M Y, eds. Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2017. 718–728. [doi: 10.18653/v1/P17-1067]
- [4] Buechel S, Hahn U. EMOBANK: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In: Lapata M, Blunsom P, Koller A, eds. Proc. of the 15th Conf. of the European Chapter of the Association for Computational Linguistics, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2017. 578–585.
- [5] Yang C, Lin KHY, Chen HH. Emotion classification using Web blog corpora. In: Yao Y, ed. Proc. of the 2007 IEEE/WIC/ACM Int'l Conf. on Web Intelligence. Washington: IEEE Computer Society, 2007. 275–278. [doi: 10.1109/WI.2007.51]
- [6] Tripathi S, Acharya S, Sharma RD, Mittal S, Bhattacharya S. Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In: Singh SP, Markovitch S, eds. Proc. of the 31st AAAI Conf. on Artificial Intelligence. Menlo Park: Association for the Advancement of Artificial Intelligence, 2017. 4746–4752.
- [7] Ekman P. Chapter 3: Basic emotions. In: Handbook of Cognition and Emotion. Hoboken: John Wiley & Sons, 1999. 45–60. [doi: 10.1002/0470013494.ch3]
- [8] Huang L, Li S, Zhou G. Emotion corpus construction on microblog text. In: Proc. of the 16th Workshop on Chinese Lexical Semantics. Heidelberg: Springer-Verlag, 2015. 204–212 (in Chinese with English abstract).
- [9] Zhu S, Li S, Chen Y, Zhou G. Corpus fusion for emotion classification. In: Matsumoto Y, Prasad R, eds. Proc. of the COLING 2016, the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka, 2016. 3287–3297.
- [10] Barrett LF. Solving the emotion paradox: Categorization and the experience of emotion. Personality and Social Psychology Review, 2006,10(1):20–46.
- [11] Duong L, Cohn T, Bird S, Cook P. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In: Zong C, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2015. 845–850.
- [12] Liang C, Berant J, Le Q, Forbus KD, Lao N. Neural symbolic machines: Learning semantic parsers on freebase with weak supervision. In: Barzilay R, Kan MY, eds. Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2017. 23–33. [doi: 10.18653/v1/P17-1003]
- [13] Wang J, Yu LC, Lai KR, Zhang X. Dimensional sentiment analysis using a regional CNN-LSTM model. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2016. 225–230.
- [14] Mishra A, Dey K, Bhattacharyya P. Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network. In: Barzilay R, Kan MY, eds. Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2017. 377–387. [doi: 10.18653/v1/P17-1035]
- [15] Cai R, Zhang X, Wang H. Bidirectional recurrent convolutional neural network for relation classification. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2016. 756–765.
- [16] Yu M, Yin W, Hasan KS, Santos C, Xiang B, Zhou B. Improved neural relation detection for knowledge base question answering. In: Barzilay R, Kan MY, eds. Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2017. 571–581. [doi: 10.18653/v1/P17-1053]
- [17] Meng F, Lu Z, Wang M, Li H, Jiang W, Liu Q. Encoding source language with convolutional neural network for machine translation. In: Zong C, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2015. 20–30.

- [18] Eriguchi A, Hashimoto K, Tsuruoka Y. Tree-to-sequence attentional neural machine translation. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2016. 823–833.
- [19] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. Advances in Neural Information Processing Systems 27 (NIPS 2014). New York, 2014. 2672–2680.
- [20] Mirza M, Osindero S. Conditional generative adversarial nets. CoRR, abs/1411.1784, 2014.
- [21] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434, 2015.
- [22] Zhang Y, Gan Z, Carin L. Generating text via adversarial training. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, eds. Proc. of the NIPS 2016 Workshop on Adversarial Training. New York, 2016. 1–6.
- [23] Zhao JJ, Kim Y, Zhang K, M.Rush A, LeCun Y. Adversarially regularized autoencoders for generating discrete structures. CoRR, abs/1706.04223, 2017.
- [24] Wu Y, Bamman D, Russell S. Adversarial training for relation extraction. In: Palmer M, Hwa R, Riedel S, eds. Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017. 1778–1783. [doi: 10.18653/v1/D17-1187]
- [25] Liu P, Qiu X, Huang X. Adversarial multi-task learning for text classification. In: Barzilay R, Kan MY, eds. Proc. of the 55th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2017. 1–10. [doi: 10.18653/v1/P17-1001]
- [26] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D, eds. Proc. of the 7th Conf. on Int'l Language Resources and Evaluation (LREC 2010). Luxembourg: European Language Resources Association, 2010. 1320–1326.
- [27] Xu G, Meng X, Wang H. Build Chinese emotion lexicons using a graph-based algorithm and multiple resources. In: Huang CR, Jurafsky D, eds. Proc. of the 23rd Int'l Conf. on Computational Linguistics (Coling 2010). Beijing: Coling 2010 Organizing Committee, 2010. 1209–1217.
- [28] Yang M, Peng B, Chen Z, Zhu D, Chow, KP. A topic model for building fine-grained domain-specific emotion lexicon. In: Toutanova K, Wu H, eds. Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2014. 421–426.
- [29] Lin KHY, Yang C, Chen HH. What emotions do news articles trigger in their readers. In: Kraaij W, Vries, AP, Clarke CLA, Fuhr N, Kando N, eds. Proc. of the 30th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: Association for Computing Machinery, 2007. 733–734. [doi: 10.1145/1277741.1277882]
- [30] Aman S, Szpakowicz S. Identifying expressions of emotion in text. In: Matoušek V, Mautner P, eds. Proc. of the 10th Int'l Conf. on Text, Speech and Dialogue. Heidelberg: Springer-Verlag, 2007. 196–205. [doi: 10.1007/978-3-540-74628-7\_27]
- [31] Mohammad SM, Turney PD. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In: Inkpen D, Strapparava C, eds. Proc. of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Stroudsburg: Association for Computational Linguistics, 2010. 26–34.
- [32] Das D, Bandyopadhyay S. Sentence level emotion tagging on blog and news corpora. Journal of Intelligent Systems, 2010,19(2): 145–162. [doi: 10.1515/JISYS.2010.19.2.145]
- [33] Bertero D, Siddique FB, Wu CS, Wan Y, Chan RHY, Fung P. Real-time speech emotion and sentiment recognition for interactive dialogue systems. In: Su J, Duh K, Carreras X, eds. Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016. 1042–1047.
- [34] Felbo B, Mislove A, Søgaard A, Rahwan I, Lehmann S. Using millions of EMOJI occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Palmer M, Hwa R, Riedel S, eds. Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017. 1615–1625. [doi: 10.18653/v1/D17-1169]



- [35] Yu LC, Wang J, Lai KR, Zhang X. Predicting valence-arousal ratings of words using a weighted graph method. In: Zong C, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2015. 788–793.
- [36] Wang J, Yu LC, Lai KR, Zhang X. Dimensional sentiment analysis using a regional CNN-LSTM model. In: Erk K, Smith NA, eds. Proc. of the 54th Annual Meeting of the Association for Computational Linguistics, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2016. 225–230.
- [37] Buechel S, Hahn U. Emotion Analysis as a regression problem-dimensional models and their implications on emotion representation and metrical evaluation. In: Kaminka GA, Fox M, Bouquet P, Hüllermeier E, Dignum V, Dignum F, Van Harmelen F, eds. Proc. of the 22nd European Conf. on Artificial Intelligence. Amsterdam: IOS Press, 2016. 1114–1122. [doi: 10.3233/978-1-61499-672-9-1114]
- [38] Preoțiuc-Pietro D, Schwartz HA, Park G, Eichstaedt JC, Kern M, Ungar L, Shulman EP. Modelling valence and arousal in facebook posts. In: Balahur A, Goot E, Vossen P, Montoyo A, eds. Proc. of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Stroudsburg: Association for Computational Linguistics, 2016. 9–15.
- [39] Yu LC, Lee LH, Hao S, Wang J, He Y, Hu J, Lai KR, Zhang X. Building Chinese affective resources in valence-arousal dimensions. In: Knight K, Nenkova A, Rambow O, eds. Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2016. 540–545.
- [40] Arjovsky M, Chintala S, Bottou L. Wasserstein gan. CoRR, abs/1701.07875, 2017.
- [41] Li Z, Zhang Y, Wei Y, Wu Y, Yang Q. End-to-end adversarial memory network for cross-domain sentiment classification. In: Sierra C, ed. Proc. of the Int'l Joint Conf. on Artificial Intelligence (IJCAI 2017). 2017. 2237–2243. [doi: 10.24963/ijcai.2017/311]
- [42] Peng M, Zhang Q, Jiang Y, Huang X. Cross-domain sentiment classification with target domain specific information. In: Gurevych I, Miyao Y, eds. Proc. of the 56th Annual Meeting of the Association for Computational Linguistics, Vol. 1. Stroudsburg: Association for Computational Linguistics, 2018. 2505–2513.
- [43] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Burges CJC, Bottou LL, Welling M, Ghahramani Z, Weinberger KQ, eds. Advances in Neural Information Processing Systems 26 (NIPS 2013). New York, 2013. 3111–3119.
- [44] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997,9(8):1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- [45] Kingma DP, Ba J. ADAM: A method for stochastic optimization. CoRR, abs/1412.6980, 2014.
- [46] Graves A. Generating sequences with recurrent neural networks. CoRR, abs/1308.0850, 2013.
- [47] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Teh YW, Titterton DM, eds. Proc. of the 13th Int'l Conf. on Artificial Intelligence and Statistics. 2010. 249–256.
- [48] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing, 2004,14(3):199–222. [doi: 10.1023/B:STCO.0000035301.49549.88]
- [49] Lamos V, Aletras N, Preoțiuc-Pietro D, Cohn T. Predicting and characterising user impact on Twitter. In: Wintner S, Goldwater S, Riezler S, eds. Proc. of the 14th Conf. of the European Chapter of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2014. 405–413.
- [50] Bitvai Z, Cohn T. Non-linear text regression with a deep convolutional neural network. In: Zong C, Strube M, eds. Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing, Vol. 2. Stroudsburg: Association for Computational Linguistics, 2015. 180–185.
- [51] Zhou C, Sun C, Liu Z, Lau FCM. A C-LSTM neural network for text classification. CoRR, abs/1511.08630, 2015.

#### 附中文参考文献:

- [2] 姚源林,王树伟,徐睿峰,等.面向微博文本的情绪标注语料库构建.中文信息学报,2014,28(5):83–91.
- [8] 黄磊,李寿山,周国栋.微博文本的情绪语料库构建.见:第 16 届汉语词汇语义学国际研讨会(CLSW 2015)会议论文集.2015. 204–212.



朱苏阳(1989-),男,江苏苏州人,博士生,主要研究领域为自然语言处理,情感分析,情绪分析.



周国栋(1967-),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为自然语言理解,机器翻译,信息抽取,信息检索,机器学习.



李寿山(1980-),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为自然语言处理,情感计算.

www.jos.org.cn

www.jos.org.cn