

## 基于情感信息融合注意力机制的抑郁症识别

陈妍<sup>1</sup>, 罗雪琴<sup>2</sup>, 梁伟<sup>1</sup>, 谢永芳<sup>3</sup>

(1. 湖南工商大学前沿交叉学院, 湖南 长沙 410205;

2. 湖南工商大学计算机学院, 湖南 长沙 410205;

3. 中南大学自动化学院, 湖南 长沙 410083)

**摘要:** 当前利用社交媒体数据预测抑郁症的研究忽视语言风格和情感随时间推移变化的特征, 缺乏对情感状态和帖子元数据特征的研究, 基于此, 提出一种基于情感信息融合注意力机制的抑郁症识别模型。首先, 在现有抑郁症识别研究的基础上, 使用文本分类卷积神经网络按时间顺序提取用户每个时间段的帖子信息和情感信息, 引入注意力机制为得到的特征矩阵分配不同的注意力权重, 进而得到用户帖子信息特征和用户情感信息特征。其次, 使用正则匹配提取文本的情感倾向信息, 拼接注意力机制输出分配权重后的情感特征矩阵, 以增强情感学习表示。然后, 加入描述社交网络帖子的元数据特征, 设计用于表示用户偏好特征的指标, 通过统计表示指标提取用户语言偏好特征。最后, 融合用户语言信息、用户情感信息、用户语言偏好信息3种特征, 建立基于多层感知机的用户抑郁状态识别模型。实验结果显示, 提出的模型精确率提高了0.051, 召回率提高了0.065, F1值提高了0.058。

**关键词:** 抑郁症识别; 注意力机制; 情感信息; 深度学习

**中图分类号:** TP389.1, TP183, TP391.1

**文献标志码:** A

**doi:** 10.11959/j.issn.2096-6652.202221

## Depression recognition based on emotional information fused with attentional mechanism

CHEN Yan<sup>1</sup>, LUO Xueqin<sup>2</sup>, LIANG Wei<sup>1</sup>, XIE Yongfang<sup>3</sup>

1. School of Advanced Interdisciplinary Studies, Hunan University of Technology and Business, Changsha 410205, China

2. School of Computer Science, Hunan University of Technology and Business, Changsha 410205, China

3. School of Automation, Central South University, Changsha 410083, China

**Abstract:** Aiming at the current research on using social media data to predict depression ignoring the characteristics of language style and emotional changes over time and lacking of research on the characteristics of emotional state and post metadata, a depression recognition model based on emotional information fused with attentional mechanism was proposed. Firstly, on the basis of the existing research on depression recognition, the text classification convolutional neural network was used to extract the post information and emotional information for each time period of the user in chronological order, and the attention mechanism was introduced to assign different attention weights to the obtained feature matrix to obtain user post information features and user emotional information features. Next, regular matching was used to extract the emotional tendency information, and the weighted emotional feature matrix output by the attention mechanism was spliced to enhance emotional learning expression. Then, metadata features describing social network posts were added, indicators that characterized user preference features were designed, and user language preference features through statistical characterization indicators were extracted. Finally, the three characteristics of user language informa-

收稿日期: 2021-11-09; 修回日期: 2021-12-17

通信作者: 谢永芳, yfxie@mail.csu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.71971080, No.71601077); 湖南省自然科学基金资助项目 (No.2020JJ4252)

**Foundation Items:** The National Natural Science Foundation of China (No.71971080, No.71601077), The Natural Science Foundation of Hunan Province (No.2020JJ4252)

tion, user emotional information, and user language preference information were combined to establish a prediction model for the user's depression state based on a multi-layer perceptron. The experimental results showed that the accuracy of the model in this paper had increased by 0.051, the recall rate had increased by 0.065, and F1 value had increased by 0.058.

**Key words:** depression recognition, attention mechanism, emotional information, deep learning

## 0 引言

现代化进程不断推进,生活压力逐渐增大,人与人之间的交流越来越少,导致抑郁症患者逐年增多。世界卫生组织 2020 年年底数据显示,全球有超 3 亿的抑郁症患者。另外,我国的抑郁症患病率超过 4.2%,并呈现低龄化趋势。抑郁症患者常表现出冲动、激越等对外攻击行为,具有自杀和扩大性自杀特点<sup>[1]</sup>。一方面,每年有近 80 万抑郁症患者自杀,严重影响了家庭幸福;另一方面,抑郁症患者他杀致人死亡的风险是未患病者的 2.6 倍<sup>[2]</sup>,对社会安全造成极大危害。此外,世界卫生组织在 2017 年的报告中预测,至 2030 年,抑郁症将高居全球疾病负担第一位,给社会经济带来沉重的负担。所以尽可能识别早期抑郁症患者,及时治疗、控制病情,成为当前社会亟须解决的问题。

大部分研究者认为工作或生活挫折导致的长期情绪低落属于正常现象,抑郁症早期患者通常无法自我察觉并及时就医,导致病情恶化。另外由于缺乏早期筛查工具,诊断方式也存在局限性,因此传统早期筛查工作面临巨大挑战。随着社交网络的发展,人们越发依赖社交媒体平台,借此表达内心感受,同时,潜在或早期的抑郁症患者习惯于通过社交媒体平台来表达情感。通过社交媒体信息识别其中患有抑郁症的用户(以下简称抑郁症用户)能有效地解决早期抑郁症患者识别的问题。通过提取用户社交网络平台留言信息,构建抑郁症识别模型,迅速抓取、追踪和锁定早期抑郁症用户。

研究者们针对社交网络中的抑郁症用户进行的识别工作是分析抑郁症潜在用户的在线行为。Park M 等人<sup>[3]</sup>利用从推特(Twitter)用户捕获的实时文本数据研究抑郁症用户情绪和语言使用规律,并在其后续工作中<sup>[4]</sup>,对活跃在社交网络平台的抑郁症用户进行了面对面采访,探讨了用户的抑郁行为特征。Xu R H 等人<sup>[5]</sup>从语言模式的角度探求

抑郁症用户的语言风格。这些研究者开创了利用社交网络识别抑郁症的方式,探讨了抑郁症用户的语言信息和行为信息,对抑郁症用户的用词规律进行了统计方面的研究,但缺乏对帖子元数据特征的分析,例如研究抑郁症用户与非抑郁症用户平均每条留言字数、字符数、字符长度等。

另外,研究者还注意到情感信息在模型学习中也是有效的。De Choudhury M 等人<sup>[6]</sup>使用回归分析法从抑郁症用户近一年的帖子中挖掘用户语言风格和情感,检测用户患重度抑郁症的潜在可能性。Shen G Y 等人<sup>[7]</sup>提出一种多模态词典学习(multimodal dictionary learning, MDL)模型检测抑郁症用户。Trotzek M 等人<sup>[8]</sup>使用文本卷积神经网络(text convolutional neural network, TextCNN)对抑郁症用户进行了分类识别。以上研究都考虑了提取抑郁症用户的情感特征,有效地提高了检测的准确性,但未考虑到抑郁症用户的情感特征在时间维度上连续变化的特点和特殊情感符号、表情符号、缩略词、大小写等在文中体现的情感信息,造成了模型情感特征缺失问题。

针对上述问题,本文提出了一种基于情感信息融合注意力机制的抑郁症识别(depression recognition based on emotional information fused with attentional mechanism, EIAM-DR)模型。融合注意力机制(attention mechanism, AM)与 TextCNN,按时间顺序提取用户每个时间段的帖子信息和情感信息。使用正则匹配提取文本中的特殊情感符号、表情符号、缩略词、大小写等,得到文本的情感倾向信息,拼接 AM 的分配权重,得到用户增强情感信息高阶特征表示。设计用于表示用户偏好特征的指标,通过统计表示指标得到用户帖子的元数据特征和用户社会性互动特征,进而得到用户语言偏好特征。最后融合用户语言信息、用户情感信息和用户语言偏好信息建立抑郁症用户识别模型。

本文后续工作组织如下:第 1 节对本文解决的问题进行了基本定义;第 2 节详细阐述了

EIAM-DR 模型；第3节介绍了实验数据集、实验参数、基准模型并详细分析了实验结果；第4节总结本文工作。

## 1 相关定义

### 1.1 问题定义

本节首先介绍一些预备知识，然后正式定义建立在社交网络平台基础上的抑郁症用户识别模型。令  $U = \{u_1, u_2, \dots, u_n\}$  表示一组社交网络平台上的用户集， $n$  表示用户的总人数。令  $D = \{d_1, d_2, \dots, d_n\}$  表示一组用户帖子集，其中  $d_i$  表示用户  $u_i$  对应的帖子集，当用户  $u_i$  在社交网络平台发表帖子时，用户的特征表示如式 (1) 所示：

$$B(u_i) = ((u_i, d_{i,0}), (u_i, d_{i,1}), \dots, (u_i, d_{i,|d_i|})) \quad (1)$$

其中，元组  $(u_i, d_{i,t})$  表示用户  $u_i$  的第  $t$  条帖子  $d_{i,t}$ ， $|d_i|$  表示用户  $u_i$  发表帖子的总量。与一般分类任务不同，本模型还需要重点关注用户在文字中的情感词、情感状态、用户语言偏好特征。令  $S = \{s_1, s_2, \dots, s_n\}$  表示一组用户情感特征， $P = \{p_1, p_2, \dots, p_n\}$  表示一组用户语言偏好特征，并拓展式 (1) 为：

$$B(u_i) = ((u_i, C_{i,0}), (u_i, C_{i,1}), \dots, (u_i, C_{i,|d_i|})) \quad (2)$$

其中，元组  $(u_i, C_{i,t})$  表示用户  $u_i$  第  $t$  条特征组  $C_{i,t}$ ，特征组  $C_{i,t} = (d_{i,t}, s_{i,t}, p_{i,t})$  是帖子  $d_{i,t}$ 、情感  $s_{i,t}$  和偏好信息  $p_{i,t}$  的集合，其中  $s_{i,t}$  表示帖子  $d_{i,t}$  对应抽取的情感信息， $p_{i,t}$  表示用户  $u_i$  的偏好信息。定义基于情感信息融合注意力机制的抑郁症识别模型为

$H(B(u_i))$ ，通过该模型预测用户是否具有患抑郁症的风险，如式 (3) 所示：

$$\hat{y}_i = H(B(u_i)) \quad (3)$$

其中， $\hat{y}_i \in \{0, 1\}$ 。当  $\hat{y}_i = 1$  时，用户  $u_i$  具有患抑郁症的风险；当  $\hat{y}_i = 0$  时，用户  $u_i$  没有患抑郁症的风险。

### 1.2 基于文本卷积神经网络的特征提取定义

在过去的几年中，深度学习被广泛应用于文本分类中，并取得了显著的效果<sup>[8-12]</sup>。由于 TextCNN<sup>[9]</sup> 具有简单和有效的特点，本文使用 TextCNN 对用户的留言或者情感进行分段卷积操作，主要由输入层、卷积层、最大池化层、隐藏层构成，并将隐藏层的输出作为下一步的输入。TextCNN 结构如图 1 所示。

### 1.3 注意力机制定义

自机器翻译<sup>[13]</sup>被提出以来，注意力机制在图像处理<sup>[14]</sup>、语音识别<sup>[15]</sup>、情感分析<sup>[16]</sup>等多种应用中得到了研究者的极大关注。上述研究<sup>[14-16]</sup>都利用了注意机制的优势。受这些研究的启发，本文将注意机制融入模型，学习情感与认知随时间变化的演化规律，具体定义如式 (4) 所示：

$$I = \text{ATTEN}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right)V \quad (4)$$

其中， $I$  表示注意力矩阵， $\text{softmax}$  表示归一化函数， $Q$  表示查询矩阵， $K$  表示键项矩阵， $V$  表示值项矩阵， $d_K$  表示  $K$  的维度。

## 2 EIAM-DR 模型

EIAM-DR 模型主要由 4 个模块构成：用户帖

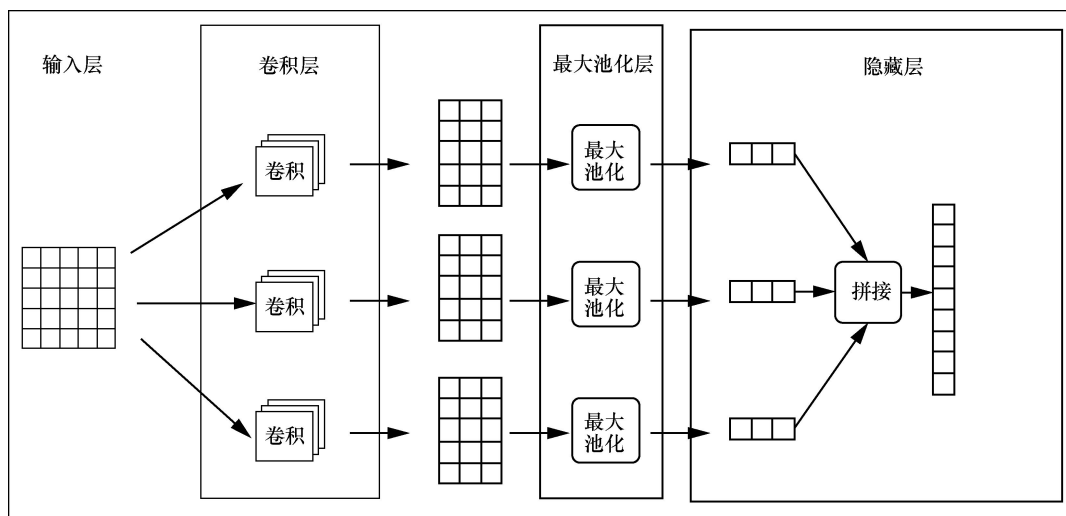


图1 TextCNN 结构

子信息提取、用户情感特征提取与增强、用户语言偏好特征提取、抑郁症用户预测分类器。假设某用户通过社交网络平台发表了一些帖子，模型首先收集该用户的帖子集，使用用户帖子信息提取模块提取用户帖子特征。再使用用户情感特征提取与增强模块学习用户在一段时间内的情感变化特征，并且学习用户每条帖子中的积极情绪、消极情绪、中性情绪以及帖子总体情感倾向的强烈程度。随后，模型进一步通过用户语言偏好特征提取模块得到用户的用词偏好特征。最后将它们的特征连接起来形成融合特征向量，使用抑郁症用户预测分类器来预测用户是否患抑郁症。EIAM-DR 模型结构如图 2 所示。

### 2.1 用户帖子信息提取

考虑到是基于社交网络平台检测用户是否患有抑郁症的，本文将基于 Twitter 数据预训练的 Glove<sup>[17]</sup>词向量作为嵌入向量，与其他平台预训练向量相比，该向量具有较好的覆盖性。首先通过文本预处理，得到降噪后的用户帖子集，将帖子  $d_{i,t}$  通过嵌入矩阵转化为字向量矩阵  $W_{ii} \in R^{m_{ii}^w \times l^w}$ ，其中  $m_{ii}^w$  是用户  $u_i$  的第  $t$  条帖子的单词个数， $w$  是帖子信息标识符， $l^w$  是单词嵌入向量的维度。本文将 TextCNN 作为特征提取器，对用户历史帖子  $d_{i,t}$  进行聚合，得到历史帖子集矩阵  $w_i' \in R^{l_{ii} \times l^w}$ ，计算过程如式 (5) 所示：

$$\begin{aligned}\tilde{w}_{ij}^h &= \sigma(f^h \cdot [W_{i,j:j+h-1}] + b^w), h=1,2,3 \\ p_{ij}^h &= \max(\tilde{w}_{i,j:j+h-1}^h), h=1,2,3 \\ w_{it}' &= p_{it}^1 \oplus p_{it}^2 \oplus p_{it}^3 \\ w_i' &= w_{i1}' \oplus w_{i2}' \oplus \dots \oplus w_{i|d_i|}'\end{aligned}\quad (5)$$

其中， $f^h$  为卷积核， $h$  为滑动窗口的高度， $b^w$  为卷积层的偏置向量， $\sigma$  为 ReLU 激活函数， $\tilde{w}_{ij}^h$  为卷积得到的第  $j$  行向量， $p_{ij}^h$  为经过池化层得到第  $j$  列的隐藏特征，拼接  $p_{it}^h$  得到帖子  $d_{i,t}$  的卷积表示  $w_{i,t}'$ 。

受参考文献[18]的启发，本文将注意机制融入 EIAM-DR 模型学习文本前后依赖关系，倾向性地关注某个时间段对解决问题起关键作用的特征，为文本信息赋予不同的注意力权重。 $w_i'$  经过注意力机制处理后得到具有前后语义关系的用户帖子的历史信息矩阵  $I_i^w \in R^{l_{ii} \times l^w}$ ，如式 (6) 所示：

$$\begin{aligned}[Q^w, K^w, V^w] &= w_i' F_{\text{linear}} \\ I^w &= \text{ATTEN}(Q^w, K^w, V^w) = \text{softmax}\left(\frac{Q^w \cdot (K^w)^T}{\sqrt{d_K^w}}\right) V^w\end{aligned}\quad (6)$$

其中， $F_{\text{linear}}$  表示线性映射函数<sup>[19]</sup>， $Q^w$ 、 $K^w$ 、 $V^w$  分别表示  $w_i'$  经过  $F_{\text{linear}}$  函数处理得到的查询矩阵、键项矩阵、值项矩阵， $d_K^w$  表示  $K^w$  的维度。

### 2.2 用户情感特征提取与增强

通过 VADER 工具<sup>[20]</sup>从帖子  $d_{i,t}$  中提取对应的情

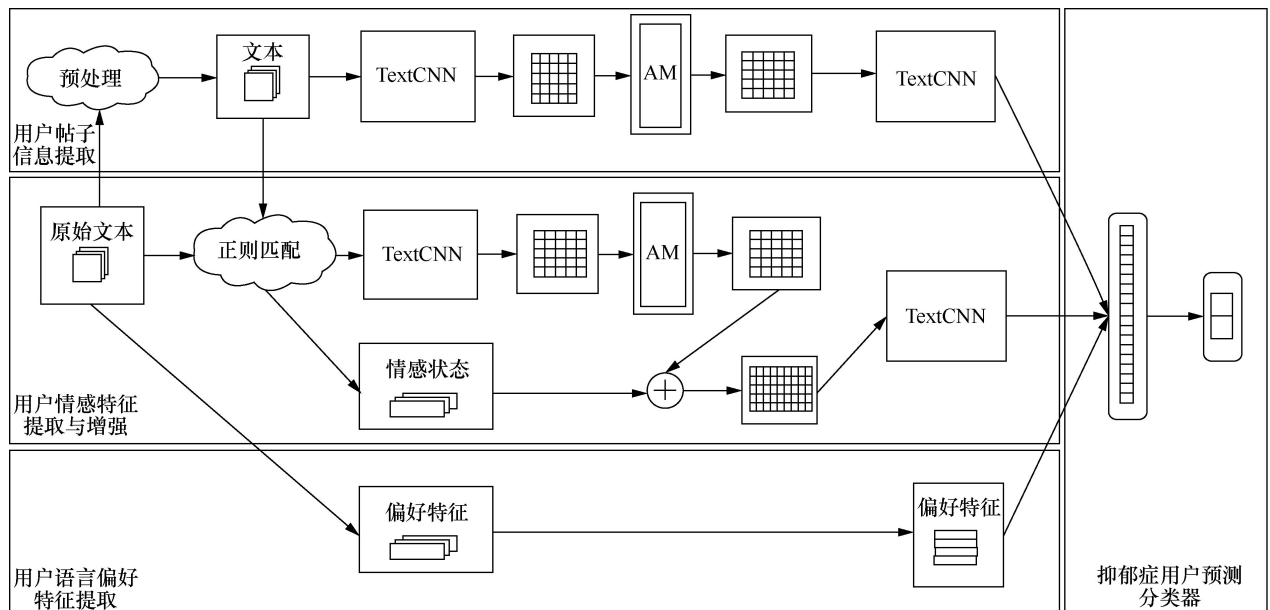


图2 EIAM-DR 模型结构

感词  $s_{i,t}$ ，使用嵌入矩阵将其转化为情感向量矩阵  $S_{it} \in \mathbb{R}^{m_{i,t}^s \times l^s}$ ，其中  $s$  为情感特征标识符， $m_{i,t}^s$  表示用户第  $t$  条帖子的情感单词总数， $l^s$  表示情感嵌入向量的维度。将对应帖子的情感词经过卷积层处理得到某条帖子情感特征  $s'_{it}$ ，如式 (7) 所示：

$$\begin{aligned}\tilde{s}_{it} &= \sigma(f \cdot [S_{it,j:j},] + b^s) \\ s'_{itj} &= \max(\tilde{s}_{it,j:j}) \\ s'_i &= s'_{i1} \oplus s'_{i2} \oplus \dots \oplus s'_{i|d_i|}\end{aligned}\quad (7)$$

其中， $s'_i$  表示用户帖子集的情感特征矩阵， $\tilde{s}_{it}$  表示情感向量矩阵  $S_{it}$  卷积后得到的矩阵， $\sigma$  为 ReLU 激活函数， $S$  表示卷积层的偏置向量， $\tilde{s}_{it}$  经过最大池化层处理得到的  $s'_{it}$  ( $s'_i \in \mathbb{R}^{|d_i| \times l^s}$ ) 表示情感集特征向量矩阵。

抑郁症用户每个时间段表达的情感会有所变化，不同时间段的情感特征对于模型的影响是不同的，因此使用注意力机制得到不同时刻特征的权重分配，给预测结果起关键作用的情感特征赋予更高的权重，由式 (8) 得到加入注意力机制之后的情感特征为  $I_i^s \in \mathbb{R}^{|d_i| \times l^s}$ 。

$$\begin{aligned}[Q^s, K^s, V^s] &= s'_i F_{\text{linear}} \\ I^s &= \text{ATTEN}(Q^s, K^s, V^s) = \text{softmax}\left(\frac{Q^s \cdot (K^s)^\top}{\sqrt{d_K}}\right) V^s\end{aligned}\quad (8)$$

其中， $Q^s$ 、 $K^s$ 、 $V^s$  分别表示  $s'_i$  经过  $F_{\text{linear}}$  函数处理得到的查询矩阵、键项矩阵、值项矩阵， $d_K$  表示  $K^s$  的维度。

为了防止模型在预处理时清理特殊情感符号、表情符号、缩略词、大小写等能表达作者情绪激烈程度的特征，本文通过 VADER 提取帖子中的情感状态  $e_i \in \mathbb{R}^{|d_i| \times k}$ ，其中  $k$  为提取的情感状态数量。拼接  $I_i^s$  和  $e_i$  得到增强的情感特征  $I_i^e \in \mathbb{R}^{|d_i| \times (l^s + k)}$ 。将  $I_i^e$  和  $I_i^w$  分别作为两个独立的卷积神经网络的输入，将文本最终向量表示  $v_i^s$  和情感特征最终向量表示  $v_i^e$  作为卷积神经网络的输出。卷积神经网络计算式如式 (9) 所示：

$$\begin{aligned}\tilde{r}_{ij}^{ch} &= \sigma(f^{ch} \cdot [I_{i,j+h-1,j}^c] + b^c), h=2,3,4 \\ p_{ij}^{ch} &= \max(\tilde{r}_{i,j:j}^{ch}), h=2,3,4 \\ v_i^c &= p_i^{c1} \oplus p_i^{c2} \oplus p_i^{c3}\end{aligned}\quad (9)$$

其中， $c$  为特征标识符， $I_i^c$  为特征类型为  $c$  的卷积层输入矩阵， $f^{ch}$  为卷积核， $h$  为卷积核的高度， $b^c$

为偏置向量， $\sigma$  为 ReLU 激活函数， $\tilde{r}$  为卷积层运算后的矩阵， $p_{ij}^{ch}$  为池化层运算后的第  $j$  列隐藏向量， $v_i^c$  为卷积网络的输出。

### 2.3 用户语言偏好特征提取

Leis A 等人<sup>[21]</sup>指出特定的语言使用模式与心理健康指标有关。语言使用模式是指个体语言表达过程中词汇使用的倾向性，可以用来区分抑郁症患者和健康个体<sup>[22]</sup>。抑郁症患者经常使用社交媒体谈论他们的疾病信息、分享治疗经验、寻求社会支持和建议、减少社交、管理他们的精神疾病<sup>[23]</sup>。因此，用户在社交网络的语言模式特征值得考虑。通过分析 Shen G Y 等人<sup>[7]</sup>贡献的数据集，并参考其分析的用户在线行为特征，本文提取了以下两个用户数据特征组：社会互动特征和帖子数据元数据。社会互动特征使用用户关注者的数量来描述用户的社会互动特征，使用用户历史帖子数量来评估用户的积极性。帖子数据元数据为描述一组数据的数据，用于分析的元特征主要包括用户平均每条帖子的单词数、用户平均每条帖子唯一单词的数量、用户平均每条帖子的停用词数量、用户平均每条帖子中链接出现的次数、用户平均每条帖子的字符数、用户平均每条帖子的提及次数。于是对提取到的用户语言偏好特征向量  $p_i$  进行归一化得到用户语言偏好特征标识符  $v_i^p$ ，如式 (10) 所示：

$$v_i^p = \frac{(p_i)_j - (p_i)_{\min}}{(p_i)_{\max} - (p_i)_{\min}} \quad (10)$$

### 2.4 抑郁症用户预测分类器

基于多层感知器模型建立抑郁症预测分类器，它的输入为由用户帖子信息特征  $v_i^w$ 、用户情感特征  $v_i^e$ 、用户语言偏好特征  $v_i^p$  等多个特征平铺级联而成的融合特征  $v_i$ ，输出为  $\hat{y}_i$ ，计算式如式 (11) 所示：

$$\begin{aligned}v_i &= v_i^w \oplus v_i^e \oplus v_i^p \\ \hat{y}_i &= \text{sigmoid}(W_m \cdot v_i + b)\end{aligned}\quad (11)$$

其中， $W_m$  和  $b$  分别为多层感知机的权重和偏置向量， $\hat{y}_i$  为第  $i$  个用户预测标签。最后，EIAM-DR 模型的损失函数如式 (12) 所示：

$$\text{loss} = -\sum [y_i \log \hat{y}_i + (-y_i) \log (1 - \hat{y}_i)] \quad (12)$$

其中， $y$  为用户真实标签， $\hat{y}$  为用户预测标签。

## 3 实验

### 3.1 实验数据

受 Benton A 等人<sup>[24]</sup>和 Coppersmith G 等人<sup>[25]</sup>

研究工作的启发, Shen G Y 等人<sup>[7]</sup>利用爬虫技术基于 Twitter 构建了一个抑郁症用户数据集 Depressed。该数据集标签满足严格的标记规则, 即用户帖子中出现“我是/我曾经是/我已经确诊了抑郁症 (I'm/I am/I was/I've been diagnosed depression)”类似的文本, 并且在一定时间内, 用户的状态符合抑郁症临床评价标准, 基于这两种情况, 将用户标记为抑郁症用户。作者进一步获取了用户在一段时间内发布的所有帖子。此外 Shen G Y 等人<sup>[7]</sup>还构建了一个非抑郁数据集 Non-Depressed, 在此数据集中, 如果用户从未发表过任何包含类似字符串“抑郁 (depress)”的帖子, 那么他们被认为是非抑郁的。数据集概述见表 1。

| 表 1 数据集概述     |       |         |
|---------------|-------|---------|
| 数据集           | 用户数/个 | 帖子数/条   |
| Depressed     | 1 402 | 209 488 |
| Non-Depressed | 2 515 | 448 038 |

### 3.2 实验细节

本文实验框架基于 Pytorch 环境, 将 Glove 模型作为词嵌入模型, 使用 Zeiler M D<sup>[26]</sup>提出的随机梯度下降法对权重和偏差进行更新, 将 Srivastava N 等人<sup>[27]</sup>提出的随机失活 (dropout) 函数应用于模型的每个隐藏层, 避免过度拟合, 并采用精确率、召回率、F1 值来评估模型的有效性。其次, 模型的参数影响模型的性能, 参数设置见表 2。其中, lr 表示训练过程的学习率, batchsize 表示每次迭代处理的数量, hiddens 表示卷积通道的数量, weight\_decay 表示权重衰减超参数, embed\_size 表示词向量的维度。

| 表 2 参数设置     |       |
|--------------|-------|
| 参数           | 值     |
| lr           | 0.003 |
| batchsize    | 20    |
| hiddens      | 100   |
| weight_decay | 0.01  |
| embed_size   | 200   |

### 3.3 对比模型

本文将以下 7 种现有的模型与 EIAM-DR 模型进行对比。

朴素贝叶斯 (naive Bayesian, NB) 模型<sup>[28]</sup>: 朴素贝叶斯算法是一种被广泛使用的分类器, 能够

将特征直接输入分类器, 并输出用户标签。

多元社交网络学习 (multiple social network learning, MSNL) 模型<sup>[29]</sup>: 对多源异构数据进行分析的多视图学习模型。

基于平滑 Wasserstein 损失的字典学习 (fast dictionary learning with a smoothed Wasserstein loss, WDL) 模型<sup>[30]</sup>: 一个字典学习模型, 它将陆地移动距离 (Wasserstein distance) 作为每个原始点及其重建点之间的拟合误差, 以利用特征共享的相似性。

多模态词典学习模型<sup>[7]</sup>: 提取了 6 个与抑郁相关的特征组, 不仅涵盖了临床抑郁症的标准, 还包括用户在社交媒体上的在线行为。

门控循环单元 (gated recurrent unit, GRU) 模型<sup>[31]</sup>: 可以解决循环神经网络中的梯度衰减问题, 并可以更好地捕捉时间序列中时间步距离较大的依赖关系。

TextCNN 模型<sup>[9]</sup>: 文本浅层特征的抽取能力强, 其中卷积操作捕捉局部信息, 池化则对局部重要的信息进行提取。

基于知识的深度对偶网络 (basic knowledge-aware deep dual network, BK-DDN) 模型<sup>[32]</sup>: 由两个神经网络分支组成的具有基本领域知识的深度对偶网络。

### 3.4 实验结果与分析

用户情感变化趋势如图 3 所示, 实验将帖子按照时间顺序排列, 研究用户的平均情感状态随时间推移的波动情况。图 3 中非抑郁症用户 (以非抑郁表示) 情感值的极差总体大于抑郁症用户 (以抑郁表示), 表明非抑郁症用户情绪变化多样, 波动较大。同时, 由图 3 可以看出抑郁症用户的消极情感值和积极情感值都显著低于非抑郁症用户, 非抑郁症用户的情感值在某一范围内波动, 抑郁症用户的消极情感值随着时间演化越来越高, 中性情感值逐渐降低。造成这种现象的原因可能是抑郁症用户容易对事件做出消极反应, 无法自我调整, 持续时间长, 而个体内心的情感可通过语言表达反映出来。长此以往, 抑郁症用户的负面词汇量逐渐增加, 消极情感呈上升趋势。

用户帖子的元数据特征如图 4 所示。从图 4 中可以看出, 抑郁症用户和非抑郁症用户的符号数、话题标签数的分布没有明显区别, 而单词个数、平均单词长度、唯一单词数、字符数、停用词数、链接数分布差别显著。与非抑郁症用户相比, 抑郁症

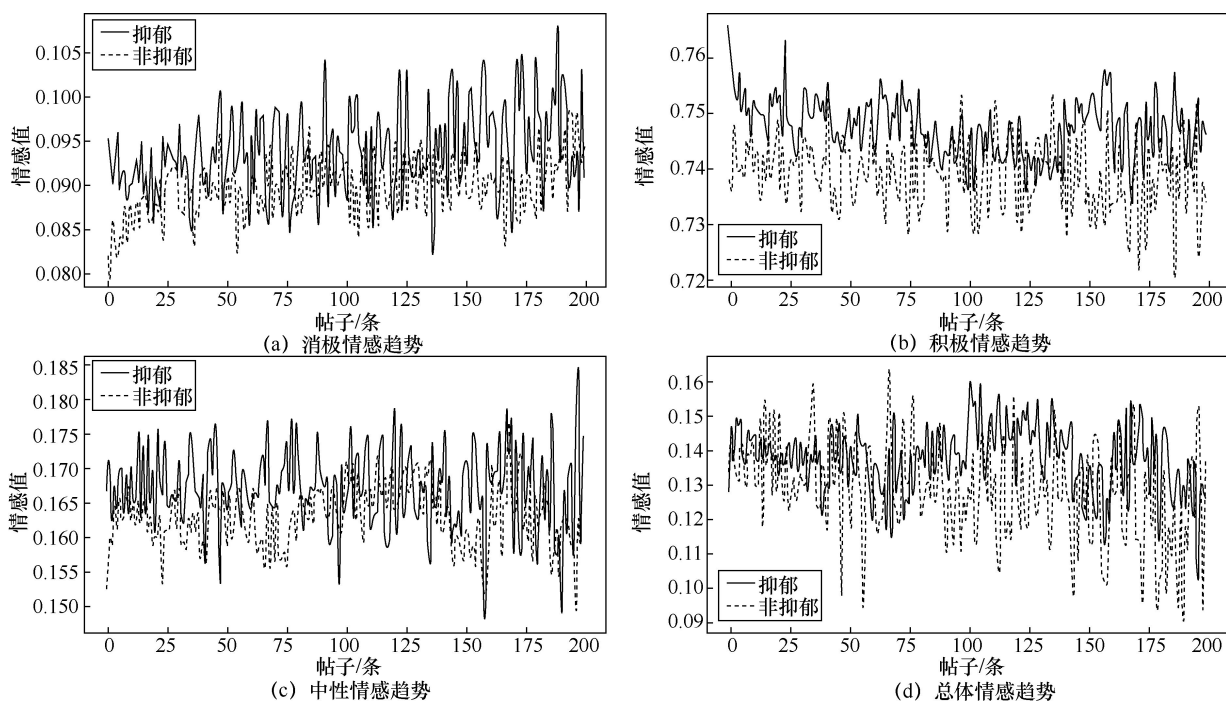


图3 用户情感变化趋势

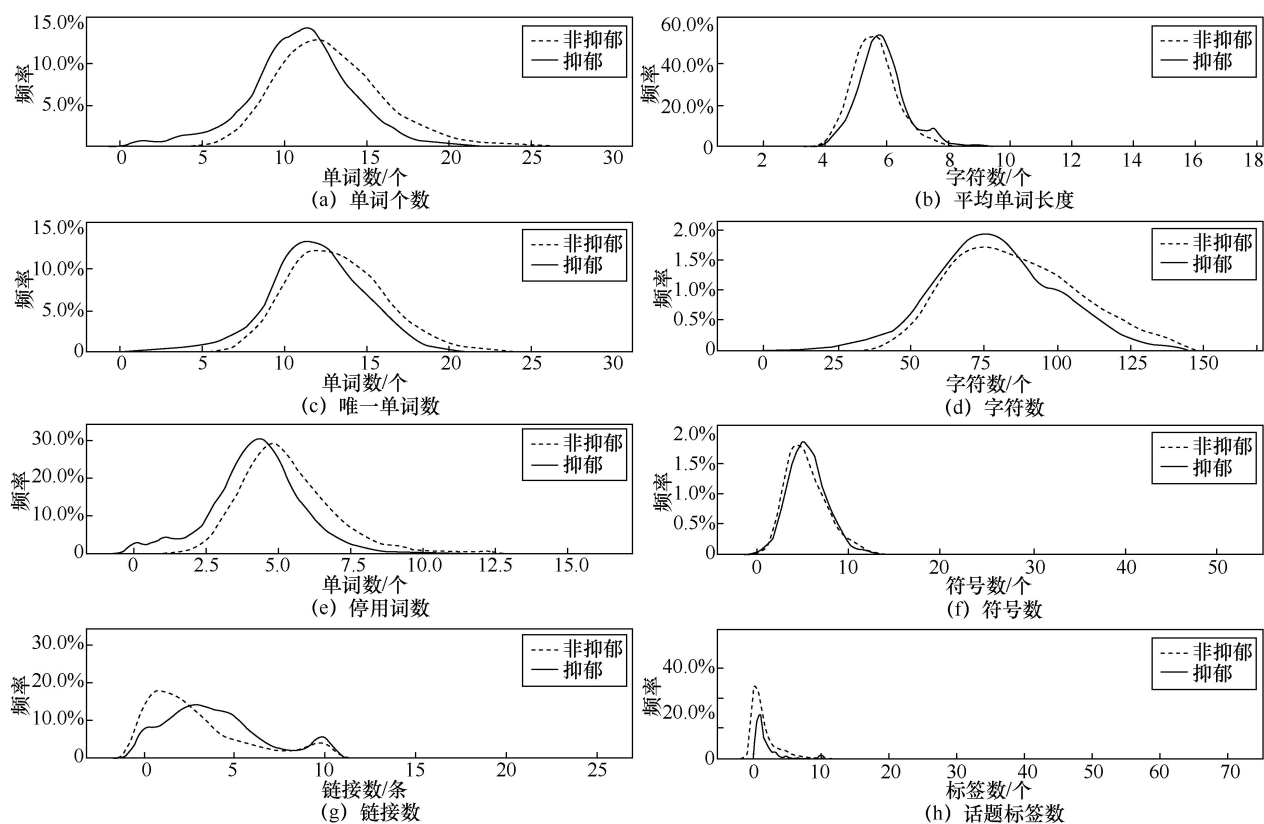


图4 用户帖子的元数据特征

用户的发文特点是用词量少、字符数多、喜欢使用停用词和链接。这些发现有助于模型识别抑郁症用户，提升模型性能。

除此之外，为了探索不同角度下 EIAM-DR 模型的效果，本文设计了3组实验，分别测试：①预训练语料对抑郁症预测分类效果的影响；②EIAM-DR 模

型与其他模型风险预测分类效果；③特征对抑郁症预测分类的实验效果的影响。

#### (1) 预训练语料对抑郁症预测分类效果的影响

Glove 预训练模型通常是在大规模数据集上训练的，词向量包含词汇语义和上下文关系信息，更具有表示意义，但是不同平台预训练学习到的深层语义特征和上下文关系有所不同，不同维度的词向量包含的语义信息也有所不同。实验比较了在维基百科 (Glove.200d) 和 Twitter (Glove.twitter.nd, nd 表示  $n$  维) 上训练的语料对抑郁症预测的实验结果，还比较了预训练词向量在不同维度对抑郁症预测实验结果的影响。

不同预训练语料在模型上测试的实验结果见表 3，表 3 中 Glove 代表全局词向量模型，其中使用 200 维的 Twitter 语料库训练的词向量的精确率为 0.918，召回率为 0.934，F1 值为 0.926，对比 Glove.200d，精确率、召回率、F1 值均有提高。实验结果证明，Twitter 语料库预训练词向量更有效。然后比较预训练的词向量维度对实验效果的影响，从表 3 中看到，精确率、召回率、F1 值随着词向量维度的增加而增大，这是因为高维度的词向量包含了更多的语义信息，因此本实验将 Glove.twitter.200d 作为模型实验的预训练向量。

表 3 不同预训练语料在模型上测试的实验结果

| 预训练语料              | 精确率   | 召回率   | F1 值  |
|--------------------|-------|-------|-------|
| Glove.200d         | 0.867 | 0.888 | 0.877 |
| Glove.twitter.25d  | 0.845 | 0.847 | 0.846 |
| Glove.twitter.50d  | 0.856 | 0.881 | 0.868 |
| Glove.twitter.100d | 0.859 | 0.895 | 0.877 |
| Glove.twitter.200d | 0.918 | 0.934 | 0.926 |

#### (2) EIAM-DR 模型与其他模型风险预测分类效果

各个模型在测试集上的抑郁症预测实验结果见表 4。TextCNN、GRU 和 NB 模型只使用了文本信息，TextCNN 与 GRU 效果相差不大，NB 的效果最差，这是由于 NB 无法学习到文字间的上文关系。WDL 的 F1 值比 NB 高 0.04，这表明潜在表示和稀疏表示在抑郁症检测中是有效的。MSNL 和 MDL 的效果都优于 WDL，验证了模态之间的建模关系有利于抑郁症用户的检测。TextCNN 在很多分类任务中都有不错的表现，于是 BK-DDN 模型搭建基于 TextCNN 的深度对偶网络，建模情感特征与文本特征，实验结果表明，BK-DDN 各个性能指标均

高于 TextCNN 模型指标，这证明了特征融合在表示学习中的重要作用。EIAM-DR 的精确率为 0.918，召回率为 0.934，F1 值为 0.926，与 BK-DDN 相比，EIAM-DR 的精确率、召回率、F1 值分别提高了 0.051、0.065、0.058。一方面，EIAM-DR 模型分段卷积用户帖子集，并通过注意力机制学习历史信息，能够有效地提取文本前后依赖关系，提高了学习更长的序列信息的能力；另一方面，从特征融合角度来看，本文模型考虑了文本信息、情感信息、用户偏好信息，与其他模型相比，实验结果均更优，这证明了融合特征的有效性。

表 4 各个模型在测试集上的抑郁症预测实验结果

| 模型      | 精确率   | 召回率   | F1 值  |
|---------|-------|-------|-------|
| NB      | 0.727 | 0.728 | 0.728 |
| MSNL    | 0.818 | 0.818 | 0.818 |
| WDL     | 0.769 | 0.768 | 0.768 |
| MDL     | 0.848 | 0.850 | 0.849 |
| GRU     | 0.825 | 0.823 | 0.824 |
| TextCNN | 0.825 | 0.852 | 0.838 |
| BK-DDN  | 0.867 | 0.869 | 0.868 |
| EIAM-DR | 0.918 | 0.934 | 0.926 |

#### (3) 特征对抑郁症预测分类的实验效果的影响

不同特征的抑郁症预测分类结果见表 5。

EIAM-DR 网络使用了所有的特征，令  $EIAM-DR_{(-p)}$  表示去掉了用户语言偏好特征的模型，令  $EIAM-DR_{(-ep)}$  表示去掉了用户语言偏好特征和情感状态的模型，令  $EIAM-DR_{(-sep)}$  表示去掉了所有的情感信息和用户语言偏好特征的模型。从表 5 中可以看出，EIAM-DR、 $EIAM-DR_{(-p)}$ 、 $EIAM-DR_{(-ep)}$ 、 $EIAM-DR_{(-sep)}$  的 F1 值分别为 0.926、0.915、0.886、0.864，从而可以得出 EIAM-DR 实验效果最好，其次是  $EIAM-DR_{(-p)}$ ，最后是  $EIAM-DR_{(-sep)}$ 。对比  $EIAM-DR_{(-sep)}$  与  $EIAM-DR_{(-ep)}$  实验结果，证明了模型对情感词的学习是有效的，对比  $EIAM-DR_{(-ep)}$  与  $EIAM-DR_{(-p)}$  的实验结果，提取特殊情感符号、表情符号、缩略词、大小写等特征对模型具有重要意义，EIAM-DR 在  $EIAM-DR_{(-p)}$  的基础上添加了用户语言偏好信息，精确率、召回率、F1 值分别比  $EIAM-DR_{(-p)}$  提升了 0.02、0.001、0.011，证明了用户语言偏好的有效性。提取的特征越多，得到的参考信息越多，模型学习更全面，所以本文提出的识

别模型对抑郁症预测分类效果最好。

表5 不同特征的抑郁症预测分类结果

| 模型                        | 精确率   | 召回率   | F1 值  |
|---------------------------|-------|-------|-------|
| EIAM-DR <sub>(-sep)</sub> | 0.853 | 0.875 | 0.864 |
| EIAM-DR <sub>(-cp)</sub>  | 0.869 | 0.904 | 0.886 |
| EIAM-DR <sub>(-p)</sub>   | 0.898 | 0.933 | 0.915 |
| EIAM-DR                   | 0.918 | 0.934 | 0.926 |

## 4 结束语

本文提出了一种用户抑郁症预测模型，用于识别活跃在社交网络的用户是否患有抑郁症。该网络模型以用户帖子为输入，从中提取文本信息，使用卷积网络对用户历史帖子建模，并通过注意力机制学习对预测结果有帮助的关键信息。提取情感信息，使用注意力机制为用户情感信息分配不同的权重以重点学习用户抑郁情绪演化特征。添加显著区分抑郁症用户和非抑郁症用户数据分布的用户语言偏好特征，并融合用户语言信息、用户情感信息、用户语言偏好信息3种特征进行用户抑郁症识别。实验结果证明，EIAM-DR模型优于本文对比模型，与其他模型相比具有更好的预测性能。该模型也可被应用于文本分析等其他任务，如社交网络中的谣言检测等。接下来笔者将探索更多的方法来优化训练过程，使其扩展到其他分类领域。

## 参考文献:

- [1] 雍那, 杜莲, 胡华, 等. 抑郁症患者攻击行为与应对方式、焦虑、抑郁的关系[J]. 第三军医大学学报, 2014, 36(11): 1133-1137.  
YONG N, DU L, HU H, et al. Correlation of aggressive behavior with coping styles, anxiety and depression in major depressed patients[J]. Journal of Third Military Medical University, 2014, 36(11): 1133-1137.
- [2] CRUMP C, SUNDQUIST K, WINKLEBY M A, et al. Mental disorders and vulnerability to homicidal death: Swedish nationwide cohort study[J]. BMJ (Clinical Research Ed), 2013, 346: f557.
- [3] PARK M, CHA C, CHA M. Depressive moods of users portrayed in Twitter[J]. Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), 2012: 1-8.
- [4] PARK M, MCDONALD D W, CHA M. Perception differences between the depressed and non-depressed users in twitter[C]//Proceedings of the 7th International AAAI Conference on Weblogs and Social Media. [S.l.:s.n.], 2013: 476-485.
- [5] XU R H, ZHANG Q P. Understanding online health groups for depression: social network and linguistic perspectives[J]. Journal of Medical Internet Research, 2016, 18(3): e63.
- [6] DE CHOUDHURY M, GAMON M, COUNTS S, et al. Predicting depression via social media[C]//Proceedings of the 7th International AAAI Conference on Weblogs and Social Media. [S.l.:s.n.], 2013: 128-137.
- [7] SHEN G Y, JIA J, NIE L Q, et al. Depression detection via harvesting social media: a multimodal dictionary learning solution[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 3838-3844.
- [8] TROTZEK M, KOITKA S, FRIEDRICH C M. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 32(3): 588-601.
- [9] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014.
- [10] LIU Y, WU Y F. FNED: a deep network for fake news early detection on social media[J]. ACM Transactions on Information Systems, 2020, 38(3): 1-33.
- [11] NASKAR D, SINGH S R, KUMAR D, et al. Emotion dynamics of public opinions on twitter[J]. ACM Transactions on Information Systems, 2020, 38(2): 1-24.
- [12] 叶俊民, 罗达雄, 陈曙. 基于短文本情感增强的在线学习者成绩预测方法[J]. 自动化学报, 2020, 46(9): 1927-1940.  
YE J M, LUO D X, CHEN S. Short-text sentiment enhanced achievement prediction method for online learners[J]. Acta Automatica Sinica, 2020, 46(9): 1927-1940.
- [13] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint, 2014, arXiv:1409.0473.
- [14] 沈继锋, 刘岳, 韦浩, 等. 基于差分特征注意力机制的无锚框多光谱行人检测算法[J]. 智能科学与技术学报, 2021, 3(3): 294-303.  
SHEN J F, LIU Y, WEI H, et al. Anchor free multispectral pedestrian detection algorithm based on differential feature attention mechanism[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 294-303.
- [15] 洪依, 孙成立, 冷严. 基于超轻量通道注意力的端对端语音增强方法[J]. 智能科学与技术学报, 2021, 3(3): 351-358.  
HONG Y, SUN C L, LENG Y. End-to-end speech enhancement based on ultra-lightweight channel attention[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 351-358.
- [16] 宋婷, 陈战伟, 杨海峰. 基于分层注意力网络的方面情感分析[J]. 大数据, 2020, 6(5): 82-91.
- [17] SONG T, CHEN Z W, YANG H F. Aspect sentiment analysis based on a hierarchical attention network[J]. Big Data Research, 2020, 6(5): 82-91.
- [18] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014: 1532-1543.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv preprint, 2017, arXiv:1706.03762v5.
- [19] 曾淦雄, 柯道. 基于3D卷积的图像序列特征提取与自注意力的车牌识别方法[J]. 智能科学与技术学报, 2021, 3(3): 268-279.

- ZENG G X, KE X. 3D convolution-based image sequence feature extraction and self-attention for license plate recognition method[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 268-279.
- [20] HUTTO C, GILBERT E. VADER: a parsimonious rule-based model for sentiment analysis of social media text[C]//Proceedings of the 8th International AAAI Conference on Web and Social Media. [S.l.:s.n.], 2014.
- [21] LEIS A, RONZANO F, MAYER M A, et al. Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis[J]. Journal of Medical Internet Research, 2019, 21(6): e14199.
- [22] HUSSAIN J, SATTI F A, AFZAL M, et al. Exploring the dominant features of social media for depression detection[J]. Journal of Information Science, 2020, 46(6): 739-759.
- [23] CAVAZOS-REHG P A, KRAUSS M J, SOWLES S, et al. A content analysis of depression-related Tweets[J]. Computers in Human Behavior, 2016, 54: 351-357.
- [24] BENTON A, COPPERSMITH G, DREDZE M. Ethical research protocols for social media health research[C]//Proceedings of the 1st ACL Workshop on Ethics in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017: 94-102.
- [25] COPPERSMITH G, DREDZE M, HARMAN C. Quantifying mental health signals in twitter[C]//Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. Stroudsburg: Association for Computational Linguistics, 2014: 51-60.
- [26] ZEILER M D. ADADELTA: an adaptive learning rate method[J]. arXiv preprint, 2012, arXiv:1212.5701.
- [27] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [28] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in Python[J]. Journal of Machine Learning Research, 2012, 12(10):2825-2830.
- [29] SONG X M, NIE L Q, ZHANG L M, et al. Multiple social network learning and its application in volunteerism tendency prediction[C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.:s.n.], 2015: 213-222.
- [30] ROLET A, CUTURI M, PEYRÉ G. Fast dictionary learning with a smoothed Wasserstein loss[C]//Proceedings of the 18th Artificial Intelligence and Statistics. [S.l.:s.n.], 2016: 630-638.
- [31] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint, 2014, arXiv:1412.3555.
- [32] LIU N, LU P, ZHANG W, et al. Knowledge-aware deep dual networks for text-based mortality prediction[C]//Proceedings of 2019 IEEE 35th International Conference on Data Engineering. Piscataway: IEEE Press, 2019: 1406-1417.

#### [作者简介]



陈妍(1981—)，女，博士，湖南工商大学前沿交叉学院副教授，主要研究方向为大数据分析、智能决策、深度学习、智慧医疗。



罗雪琴(1996—)，女，湖南工商大学计算机学院硕士生，主要研究方向为自然语言处理、深度学习、智慧医疗。



梁伟(1982—)，男，博士，湖南工商大学前沿交叉学院副教授，主要研究方向为自然语言处理、大数据分析、深度学习、智慧医疗。



谢永芳(1972—)，男，博士，中南大学自动化学院教授、博士生导师，主要研究方向为深度学习、智能优化、智能控制、知识自动化。