

基于知识蒸馏和评论时间的文本情感分类新方法

王友卫¹, 刘 奥¹, 凤丽洲²

(1. 中央财经大学信息学院, 北京 100081; 2. 天津财经大学统计学院, 天津 300222)

摘要:针对现有的情感分类方法普遍未能充分考虑用户个性化特征且忽略时间因素对情感分类结果的影响的问题,提出一种基于知识蒸馏和评论时间的文本情感分类新方法。首先,为解决数据集中高质量标注数据较少的问题,采用 RoFormer-Sim 生成模型对训练文本数据增强;然后,引入评论时间属性,从用户历史评论中提取用户的个性化信息,提出基于多特征融合的评论文本情感得分预测模型;最后,为提高针对冷启动用户的泛化性能,引入知识蒸馏理论,利用 SKEP 模型对基于多特征融合的情感分类模型进行通用性增强。在从中文股吧爬取的真实数据集上的实验结果表明:与 SKEP、ELECTRA 等典型方法相比,本文方法在准确率上分别提高了 3.1% 和 0.9%,在 F_1 值上分别提高了 2.7% 和 1.0%,验证了其在改善情感分类表现方面的有效性。

关键词: 计算机应用;情感分类;知识蒸馏;数据增强;历史评论

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1671-5497(2025)05-1664-11

DOI: 10.13229/j.cnki.jdxbgxb.20230845

New method for text sentiment classification based on knowledge distillation and comment time

WANG You-wei¹, LIU Ao¹, FENG Li-zhou²

(1. School of Information, Central University of Finance and Economics, Beijing 100081, China; 2. School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China)

Abstract: Aiming at the problem that the existing sentiment classification methods generally did not fully consider the user's personalized characteristics and ignore the influence of time factor on the sentiment classification results, a new method for text sentiment classification based on knowledge distillation and comment time is proposed. Firstly, in order to solve the problem of less labeled data with high quality, the RoFormer-Sim generative model is used to augment the training text data. Then, a sentiment score prediction model of comment text based on multi-feature fusion is proposed by introducing comment time attribute to extract users' personalized information from user's historical comments. Finally, in order to

收稿日期: 2023-08-09.

基金项目: 中央财经大学科研创新团队支持计划项目(202516);中央财经大学新兴交叉学科建设项目(202106);天津市教委科研项目(2023SK115);国家社科一般项目(20BTJ058).

作者简介: 王友卫(1987-),男,副教授,博士.研究方向:内容安全,深度学习,数据挖掘.

E-mail: ywwang15@126.com

improve the generalization performance for cold start users, the knowledge distillation theory is introduced, and SKEP model is used to enhance the versatility of the sentiment classification model based on multi-feature fusion. The experimental results on the real dataset crawled from the Chinese stock page show that compared with typical methods such as SKEP and ELECTRA, the accuracy of the proposed method is improved by 3.1% and 0.9%, and the F_1 value is increased by 2.7% and 1.0%, respectively, which verifies its effectiveness in improving the performance of sentiment classification.

Key words: computer application; sentiment classification; knowledge distillation; data augmentation; historical comment

0 引言

随着互联网技术的快速发展,股吧平台成为越来越多投资者进行信息交流和获取的重要平台。通过针对股吧评论的情感分析,可以帮助投资者了解市场动向,提供给投资者相应的投资建议,也可以在一定程度上预测市场波动,帮助企业提前做好应对措施以化解风险。近年来,深度学习因其卓越的分类表现,被更多地应用在情感分类领域。Umer等^[1]将极度随机树(Extra tree)与卷积神经网络(Convolutional neural networks, CNN)两个模型单独训练,再通过软投票机制将模型集成,以提高模型的准确率和稳健性。Lan等^[2]提出了ALBERT(A lite BERT)预训练模型,该模型减少了BERT模型整体的参数量,加快了训练速度。但是,该模型的下游微调任务仍需要大量文本数据,因此,针对数据匮乏的中文金融领域无法获得较好的情感分类性能。Liu等^[3]提出了FinBERT(BERT for financial text mining)预训练模型,该模型在BERT模型的基础上通过大规模英文金融语料训练得到,解决了英文金融领域缺乏预训练模型的问题,但因训练语料的限制,该模型无法应用在中文金融评论领域。赵亚欧等^[4]提出了基于ELMo(Embeddings from language models)和Transformer的混合模型,并将长短时记忆网络(Long short-term Memory, LSTM)和多头注意力机制引入模型,解决了评论文本双向语义和一词多义的问题。但是,该方法仅提取评论文本自身的情感特征,未能准确挖掘用户个性特征和社会关系对文本情感分析的影响。为此,Yang等^[5]利用用户与用户、用户与博文之间的关系构建图,并利用大规模信息网络嵌入算法(Large-scale information network embedding, LINE)得到含有社交关系的节点向量。但

是,以上方法未能对用户丰富的历史评论信息进行深入挖掘。为此,蒋宗礼等^[6]利用分层的多头注意力机制从多个角度挖掘用户和产品信息,更全面地捕捉用户与产品信息对情感分类的影响,取得了较好的实验结果。

研究发现,目前的情感分类方法仍面临以下问题:①金融领域质量合格的标注评论较为缺乏,限制了情感分类模型的准确性和泛化性;②未能考虑用户属性特征、用户历史评论特征、用户社交关系特征的综合影响,忽略了用户个性化特征和时间属性对情感分类的作用;③情感分类结果依赖于有限的训练数据,模型针对其他相关领域的迁移性和通用性较差。为此,本文提出一种基于知识蒸馏和评论时间的文本情感分类新方法。该方法首先利用RoFormer-Sim生成模型对评论文本进行数据增强;然后,利用用户大量无标注的历史评论提取用户个性特征,并通过引入时间遗忘函数以充分考虑时间因素对情感分类结果的影响;最后,提出基于用户相似度的冷启动用户表示方法,并引入蒸馏学习理论提高模型针对不同领域的泛化性能。

1 相关理论

1.1 知识蒸馏

知识蒸馏^[7]是一种有效的模型压缩方法。它能够将复杂的教师模型所学习到的知识迁移到轻量级的学生模型中。蒸馏训练主要是将复杂教师模型生成的软标签作为监督信号加入学生模型的训练中,以完成知识的迁移,达到提高学生模型分类准确率和泛化性的目的^[8]。教师模型的软标签通过引入温度系数 T 以对模型的输出进行平滑处理而得到,温度系数 T 的引入使软标签分布更加平滑,学生模型学到的知识也就更加丰富。软标签的计算公式如下:

$$y_i^T = \text{Soft max} - T(z_i) = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (1)$$

式中: y_{ij}^T 为软标签; z_i, z_j 为模型输出; T 为温度系数;Softmax- T 函数为引入温度系数 T 的Softmax函数。

整个蒸馏学习的损失 L_{KD} 为学生模型的交叉熵损失 L_{CE} 和蒸馏损失 L_{DL} 的加权求和,计算公式为:

$$L_{KD} = aL_{CE}(p, y) + (1 - a)L_{DL}(p^T, q^T)T^2 \quad (2)$$

式中: $L_{CE}(p, y)$ 为学生模型的硬标签(Softmax函数输出) p 与真实标签 y 的交叉熵损失; $L_{DL}(p^T, q^T)$ 为学生模型的软标签 p^T 与教师模型的软标签 q^T 的KL散度损失; a 为权重因子;乘以 T^2 是为了均衡这两个损失。

1.2 情感知识增强预训练模型(Sentiment knowledge enhanced pre-training for sentiment analysis, SKEP)

SKEP模型^[9]是基于情感知识增强的情感预训练模型。SKEP模型首先使用点互信息(Pointwise mutual information, PMI)^[10]的方法对大量无标记数据自动挖掘情感知识,包括情感词、情感词极性以及观点搭配;然后,通过以下方式对原始句子的部分情感词语进行屏蔽:随机遮掩最多两对观点搭配、随机遮掩不超过当前句子token总数10%的情感词、当遮掩的比例不足10%时随机补充遮掩通用词至10%;最后,该模型将复原被屏蔽的情感信息。鉴于SKEP模型准确性高、通用性好、一键式预测等优势,本文将结合蒸馏学习思想利用SKEP模型提高情感分类模型的泛化性能。

2 问题描述

定义1 用户当前评论:当前进行情感分类的用户评论文本,表示为 $U_{ij}^C(1 \leq i \leq N)$,代表用户 i 的第 j 条当前评论。

定义2 用户历史评论集:用户一段时间内在平台上发表的所有历史评论,定义为 $U_i^P = \{U_i^P\}(1 \leq i \leq N)$,其中 $U_i^P = \{U_{ij}^P\}(1 \leq j \leq C_i, C_i$ 为用户 i 具有的历史评论数), U_{ij}^P 表示第 i 个用户的第 j 条历史评论。

定义3 社会关系:用户间社会关系表示为 $G = (V, A)$,其中 $V = \{v_i\}(1 \leq i \leq N, |V| = N)$ 为用户集合, $A \in \mathbb{R}^{N \times N}$ 为用户的邻接矩阵,能够表达有

向图 G 的结构信息, A_{ij} 的定义为:

$$A_{ij} = \begin{cases} 1, & \text{用户节点 } v_i \text{ 关注用户节点 } v_j \\ 0, & \text{其他} \end{cases} \quad (3)$$

定义4 用户属性特征:用户 i 的属性特征为 F_i ,其中, $F_i = \{fun_{ij}\}(1 \leq j \leq n, n$ 为用户属性个数), fun_{ij} 为用户 i 的第 j 个属性值,通常指用户的粉丝数、关注人数、影响力值等信息。

在此基础上,本文综合多种特征的影响,将情感分类函数定义为 $F(h_{ij}, \Omega, S)$ 。若 $F(h_{ij}, \Omega, S) = 1$,则 U_{ij}^C 的情感类别为积极,否则,其情感类别为消极。其中, h_{ij} 为 U_{ij}^C 对应的多特征融合后的向量, Ω 为模型所需参数, S 为知识蒸馏教师模型。为便于理解,本文其余相关符号的定义如表1所示。

表1 本文主要符号及含义

Table 1 Main symbols and meanings of this article

符号	含义
$word_k^j$	将 U_{ij}^P 分词、去停用词等操作后的第 k 个词
$h_{jk}^W \in \mathbb{R}^m$	$word_k^j$ 对应的词向量
$h_{ij}^S \in \mathbb{R}^m$	U_{ij}^P 的句向量
$h_i^H \in \mathbb{R}^m$	用户 i 的历史评论向量
$h_i^F \in \mathbb{R}^{m+n}$	用户 i 的特征向量
$h_i^A \in \mathbb{R}^n$	用户 i 的属性向量
$h_{ij}^C \in \mathbb{R}^{m+n}$	用户当前评论 U_{ij}^C 的句向量
$h_i^R \in \mathbb{R}^{m+n}$	用户 i 的关系向量
N	用户总数
m	词向量的维数

3 基于知识蒸馏和评论时间的文本情感分类新方法

本文方法主要分为两个模块:基于多特征融合的评论文本情感得分预测模型与基于知识蒸馏和用户相似性的模型通用性增强,具体过程如图1所示。

3.1 基于多特征融合的评论文本情感得分预测模型(MFM)

3.1.1 评论文本数据增强

股吧的评论文本有较强的不规范性,充斥着大量的网络用词、口语词,同时有一定的专业性,包含较多投资专业术语。这导致文本的噪声较大,质量合格的标注数据较少。为此,本文采用RoFormer-Sim模型^[11]对评论文本进行数据增强。RoFormer-Sim是同时具备相似句生成和相似句检索能力的生成式语言模型。该模型在

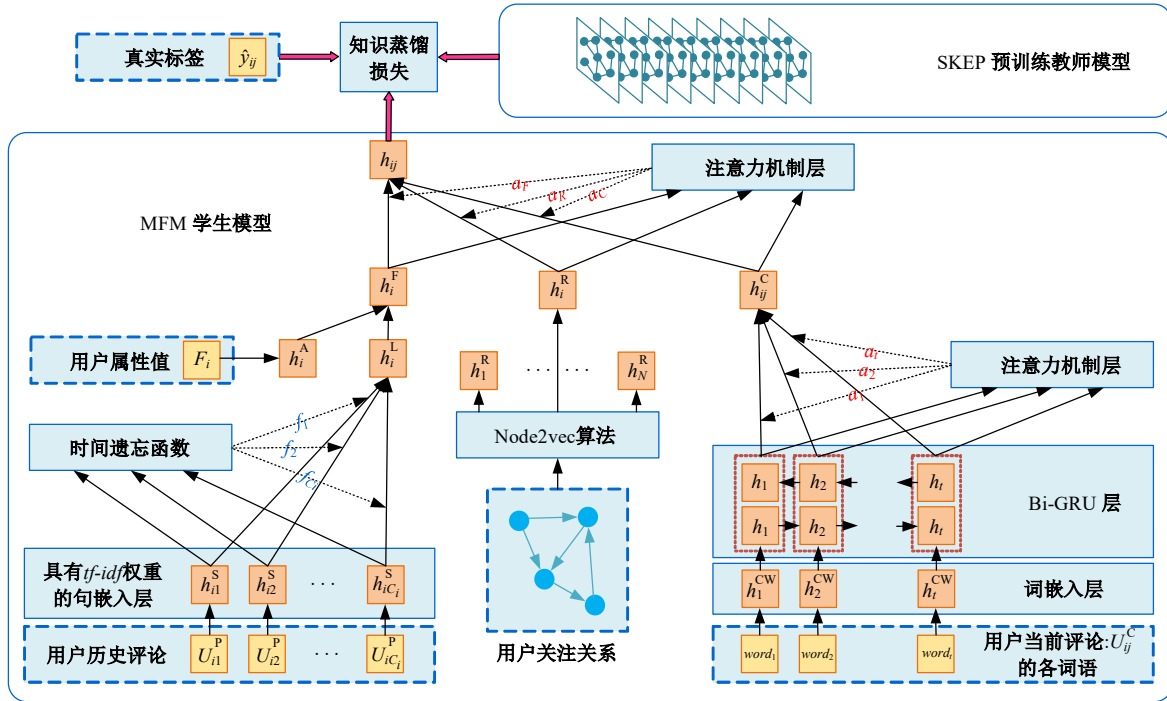


图 1 本文方法整体框架

Fig. 1 Overall framework of the proposed method

SimBERT^[12]的基础上,增加了仿 BART(Bidirectional and auto-regressive transformers)^[13]式训练、模型蒸馏以及一般句式语料的训练,是对 SimBERT 相关技术进一步的整合和优化。

3.1.2 用户特征向量生成

为了有效考虑情感分类任务里评论的时间属性,本文在为每个用户构建个性化用户特征向量时,引入可解释性更强的时间遗忘函数,获得包括历史评论中蕴含的个性信息和粉丝数、访问量、影响力等属性信息,以此融合了用户的用词习惯、性格特点、态度偏好、遗忘属性、影响力及活跃程度等个性化因素。具体过程如下:

为获得用户 i 的第 j 个历史评论 U_{ij}^p 的句向量 h_{ij}^s ,将 U_{ij}^p 所有词对应的词向量乘以该词相应的 $tf-idf$ 值,以体现句子中每个词的重要程度,计算公式为:

$$h_{ij}^s = \frac{\sum_{k=1}^{t_{ij}^p} h_{ijk}^w \times \exp(T_k)}{t_{ij}^p} \quad (4)$$

式中: $\exp()$ 表示以 e 为底的指数函数; T_k 为词语 $word_k^j$ 的 $tf-idf$ 值,其计算方法为:

$$T_k = tf_k \times idf \quad (5)$$

$$tf_k = \frac{n_k}{t} \quad (6)$$

$$idf = \lg \frac{n_p}{n_{pk} + 1} \quad (7)$$

式中: tf_k 为 $word_k^j$ 的词频; idf 为逆向文件频率; n_k 为词 $word_k^j$ 在本评论中出现的次数; n_p 为所有评论的个数; n_{pk} 为包含词 $word_k^j$ 的评论数。

可见, T_k 越大,说明 $word_k^j$ 越重要,对句向量 h_{ij}^s 的影响越大。

人们的情绪特征和态度偏好通常会随着时间的推移而发生较大的改变,德国心理学家 Ebbinghaus^[14] 研究发现了艾宾浩斯遗忘曲线,揭示了人们对事物的遗忘过程呈现非线性的变化。根据此规律,构造与艾宾浩斯遗忘曲线类似的基于时间的遗忘函数,定义为:

$$f(u, i) = e^{-a \times \frac{t_i - t_0}{T_u}} \quad (8)$$

式中: $f(u, i)$ 为用户 u 的第 i 个历史评论的时间权重,其数值在 $(0, 1)$ 之间; t_i 为用户 u 的第 i 个历史评论所发表的时间; t_0 为用户最早发表评论的时间; T_u 为用户最早发表评论与最近发表评论的时间间隔; a 为权重调控因子, $a \in [0, 1]$ 。

进一步地,计算得到用户 i 的历史评论向量 h_i^l ($1 \leq i \leq N$, $h_i^l \in \mathbb{R}^m$, m 为词向量维数),公式为:

$$h_i^l = \frac{\sum_{j=0}^{C_i} f(i, j) \times h_{ij}^s}{C_i} \quad (9)$$

式中: C_i 为用户 i 具有的历史评论数; $f(i, j)$ 为 U_{ij}^p 对应的时间遗忘函数值。

按照定义 4 提取用户 i 的属性集 F_i 所对应的属性值, 为避免不同维度跨度大小的影响, 对该向量每维数据进行 Max-Min 归一化操作, 以得到用户 i 的属性向量 $\mathbf{h}_i^A \in \mathbb{R}^n$ 。在此基础上, 将其与用户历史评论向量 \mathbf{h}_i^L 拼接, 得到用户 i 的特征向量 $\mathbf{h}_i^F \in \mathbb{R}^{m+n}$, 公式为:

$$\mathbf{h}_i^F = \mathbf{h}_i^L \parallel \mathbf{h}_i^A \quad (10)$$

式中: \parallel 为向量拼接操作。

3.1.3 用户社会关系向量生成

为提高用户社会关系学习效果, 本文利用 node2vec 算法^[15] 对用户社会关系网络结构进行学习, 通过控制节点序列产生过程中的游走趋向将更大范围的网络结构信息体现到节点向量中。使用 node2vec 算法可得到一系列节点序列, 并通过 skip-gram 模型训练获得用户社会关系向量 $H_R = [\mathbf{h}_i^R] (1 \leq i \leq N, \mathbf{h}_i^R \in \mathbb{R}^{m+n})$, 对应的目标函数为:

$$L = \frac{1}{N} \sum_{i=1}^N \sum_{-c \leq j \leq c, j \neq 0} \log p(\omega_{i+j} | \omega_i) \quad (11)$$

式中: N 为用户数量; c 为窗口大小; ω_i 为第 t 个用户节点。

$$p(\omega_{i+j} | \omega_i) = \frac{\mathbf{u}_{\omega_{i+j}}^T \mathbf{v}_{\omega_i}}{\sum_{i=1}^{|\mathcal{V}|} \exp(\mathbf{u}_{\omega_i}^T \mathbf{v}_{\omega_i})} \quad (12)$$

式中: \mathbf{u}_w 与 \mathbf{v}_w 分别为节点 w 是背景节点和中心节点的向量表达。

3.1.4 基于多特征融合的情感分类

本文选取字长不超过 400 的用户评论进行标注, 形成带有标注的用户当前评论数据 $U_{ij}^c (1 \leq i \leq N)$, 对应的标签为 $\hat{y}_{ij} (\hat{y}_{ij} \in \{0, 1\})$ 。进一步地, 对当前评论进行分词、去停用词等操作, 并利用预训练的 skip-gram 模型将各词语转换为词向量, 得到词向量矩阵 H_{ij}^{CW} (该矩阵由词向量 \mathbf{h}_{ij}^{CW} 组成, $1 \leq t \leq t_{ij}^C, t_{ij}^C$ 为 H_{ij}^{CW} 所含词语个数), 为了充分利用上下文信息, 保证同时对数据上下文的处理能力, 引入双向 GRU 单元^[16] (BiGRU), 将 H_{ij}^{CW} 中的节点向量依次输入 BiGRU 中, 以获得所有词向量的矩阵 H_{ij} (该矩阵由 \mathbf{h}_t^B 组成, \mathbf{h}_t^B 为 BiGRU 网络输出的第 t 个词语向量, $1 \leq t \leq t_{ij}^C$), 其中 \mathbf{h}_t^B 的计算公式为:

$$\vec{\mathbf{h}}_t = \overrightarrow{\text{GRU}}(\mathbf{h}_{ij}^{CW}, \mathbf{h}_{t-1}) \quad (13)$$

$$\overleftarrow{\mathbf{h}}_t = \overleftarrow{\text{GRU}}(\mathbf{h}_{ij}^{CW}, \mathbf{h}_{t+1}) \quad (14)$$

$$\mathbf{h}_t^B = [\vec{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t] \quad (15)$$

式中: $\vec{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t$ 分别为 GRU 网络从前往后、从后往前处理词语序列得到的第 t 个词语向量; \mathbf{h}_t^B 由 $\vec{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t$ 拼接得到; $[\ , \]$ 为向量拼接符。

为了识别不同词对情感分类的重要程度, 将 BiGRU 的输出 H_{ij} 输入注意力层中, 可将用户当前评论 U_{ij}^c 转化为相应的句向量 $\mathbf{h}_{ij}^C \in \mathbb{R}^{m+n}$, 公式为:

$$\mathbf{h}_{ij}^C = \text{Attention}(H_{ij}) \quad (16)$$

式中: Attention() 为采用一层全连接神经网络的注意力层, 公式为:

$$A = \text{Softmax}(\mathbf{W}_A \mathbf{M} + \mathbf{b}_A) \quad (17)$$

$$\mathbf{M}' = \mathbf{M} \odot A \quad (18)$$

$$\mathbf{z} = \sum_i \mathbf{M}'_i \quad (19)$$

式中: \mathbf{M} 为注意力机制的输入矩阵; $\mathbf{W}_A, \mathbf{b}_A$ 分别为权重矩阵和偏置矩阵; A 为注意力分数矩阵, 形状与 \mathbf{M} 相同; \mathbf{M}' 为中间向量矩阵; \mathbf{z} 为注意力机制的输出, \odot 表示矩阵哈达玛积。

在此基础上, 利用与式 (16) 相同的注意力机制以获得当前评论文本的最终向量表达: $\mathbf{h}_{ij} \in \mathbb{R}^{m+n}$, 公式为:

$$\mathbf{h}_{ij} = \text{Attention}([\mathbf{h}_{ij}^C, \mathbf{h}_i^F, \mathbf{h}_i^R]) \quad (20)$$

进一步地, 将 \mathbf{h}_{ij} 输入全连接层中, 以获得当前评论 U_{ij}^c 所对应的情感得分 $s_{ij} \in \mathbb{R}^2$, 公式为:

$$s_{ij} = \mathbf{W}_0 \mathbf{h}_{ij} + \mathbf{b}_0 \quad (21)$$

式中: $\mathbf{W}_0 \in \mathbb{R}^{2 \times (m+n)}, \mathbf{b}_0 \in \mathbb{R}^2$ 为训练权重矩阵。

现给出本文基于多特征融合的评论文本情感得分预测模型 (MFM) 的执行过程, 如算法 1 所示。

算法 1 基于多特征融合的评论文本情感得分预测

输入: 用户 i 的当前评论数据集 $U_i^c = \{U_{ij}^c\} (1 \leq j \leq R_i, R_i$ 为用户 i 具有的当前评论数), 用户关注关系 $G = (V, A)$, 用户 i 的历史评论集 $U_i^p = \{U_{ij}^p\} (1 \leq j \leq R_i)$ 和用户 i 的属性集 F_i

输出: U_i^c 的情感得分向量 s_i

1. 对 F_i 进行 Max-Min 处理, 得到用户 i 的属性向量 $\mathbf{h}_i^A \in \mathbb{R}^n$
2. 利用 node2vec 算法处理用户社会关系图 $G = (V, A)$, 得到用户的社会关系向量: $H_R = [\mathbf{h}_i^R] (1 \leq i \leq N)$
3. for all $U_{ij}^p \in U_i^p$ do
4. 利用式 (4) 获得 U_{ij}^p 对应句向量 \mathbf{h}_{ij}^S
5. 利用式 (8)(9) 获得用户 i 的历史评论向量 \mathbf{h}_i^L
6. end for

7. 利用式(10)获得用户*i*的特征向量 h_i^f
8. for all $U_{ij}^c \in U_i^c$ do
9. 将 U_{ij}^c 所有词对应的词向量输入 BiGRU 网络中, 并利用式(16)获得句向量 h_{ij}^c
10. 利用式(20)融合 h_i^f 、 h_{ij}^c 和 h_{ij}^r , 得到 U_{ij}^c 对应的特征向量 h_{ij}
11. end for
12. 将 $H_i = [h_{ij}] (1 \leq j \leq R_i)$ 输入全连接层, 利用式(21)获得情感得分向量 $s_i = [s_{ij}] (1 \leq j \leq R_i, s_{ij}$ 为 U_{ij}^c 的情感得分)

3.2 基于知识蒸馏和用户相似性的模型通用性增强

3.2.1 冷启动用户表示

为了提高模型针对训练集中未出现用户(冷启动用户)的表示能力,本文提出针对冷启动用户的关系向量表示方法。具体而言,首先,通过遍历数据集中的所有 N 个用户,计算冷启动用户 v 的特征向量与现有全部用户 i 的余弦相似度,如式(22)所示;然后,将所有用户按余弦相似度值大小降序排序,得到相似度值排前5的用户;最后,对这5个用户的关系向量取平均值,可得到冷启动用户 v 的关系向量 $h_v^R \in \mathbb{R}^{m+n}$,如式(23)所示:

$$\cos(v, i) = \frac{h_i^F \times h_v^F}{\|h_i^F\| \|h_v^F\|} \quad (22)$$

$$h_v^R = \frac{\sum_{i \in \text{top5_list}} h_i^R}{5} \quad (23)$$

式中: $\cos(v, i)$ 为冷启动用户 v 与用户 i 的余弦相似度; top5_list 为余弦相似度值前5的用户集。

3.2.2 模型通用性增强

单纯使用基于多特征融合的情感分类方法易造成模型对训练集数据依赖性强的问题,使模型预测结果针对特定文本有一定的偏向性。为此,本文引入知识蒸馏理论,以通用情感分类模型 SKEP 为教师模型,以多特征模型 MFM 为学生模型,将 SKEP 模型的通用性情感分类知识泛化到

MFM 上,以此提高情感分类模型对其他评论领域的通用性和准确率。为避免过拟合,在蒸馏训练的损失函数中加入了 L_2 正则化惩罚项,计算公式为:

$$L_{KD} = aL_{CE} + (1-a)T^2L_{DL} + \lambda\|\theta\|^2 \quad (24)$$

式中: L_{CE} 、 L_{DL} 分别为学生模型的交叉熵损失与蒸馏损失; a 为权重因子; T 为温度系数; θ 为模型全部参数; λ 为 L_2 正则化的惩罚系数, λ 越大,惩罚力度越大。

$$L_{CE} = -\sum_{i=1}^N \sum_{j=1}^{R_i} \hat{y}_{ij} \ln p_{ij}, L_{DL} = \sum_{i=1}^N \sum_{j=1}^{R_i} q_{ij}^T \ln \left(\frac{q_{ij}^T}{p_{ij}^T} \right) \quad (25)$$

式中: \hat{y}_{ij} 为 U_{ij}^c 的真实情感标签; p_{ij} 为 MFM 模型的硬标签; p_{ij}^T 为 MFM 模型的软标签; q_{ij}^T 为 SKEP 模型的软标签。

4 实验结果与分析

4.1 数据集

针对东方财富网股吧平台的用户,使用八爪鱼专业版软件爬取其在2021年11月之前发布的评论(最大不超过60条)。每个用户选取最多20个关注者,最终得到了用户评论数据集,数据集中共含有437个用户,17391条评论,479条关注边数。对每个用户爬取包括粉丝数、吧龄、访问量、关注的人数、影响力值等属性信息。在所有的用户评论中,选取标题和正文总字数不超过400的评论进行积极性和消极性标注,以此得到5791条标注数据。数据集中部分评论示例如表2所示。

为了验证用户历史评论对情感分类结果的影响,本文选用发表时间为最近一周的标注评论为测试集,除去测试集后,训练集、验证集分别占剩余数据的7/8与1/8。实验使用准确率(Acc)、 F_1 值^[17]和均方根误差(Root mean square error, RMSE)^[18]来衡量模型的性能,每次实验的结果通过10次随机实验的平均值得到。实验软件

表2 用户评论示例

Table 2 Examples of user's comment

用户名	标题+正文	情感标记	发表时间
股友3663E5w329	储能最差的股!盛洪是储能板块最差的,前面走过几倍了,是不可能拉升	0	2021/09/17
芝加哥夜鹰	这个股价是不可能回头了,唉,郁闷呢。大帝一出手,股市抖三抖	0	2021/09/11
未来可期30721	涨停无疑了,连板股,吃大块[微笑]	1	2021/05/26
漫步冥王星	塔牌这一跌,跌得没头没脑。另一方面来看,很好的进货机会	1	2021/04/28
松松雪7856	已经大举加仓。锁仓!关电脑	1	2021/04/2
喉喉喉喉369	这货踏踏实实跌,主力底下等着呐,黑庄,这股就是笑话	0	2021/02/1

环境为:Python 3.7.0+TensorFlow 2.1.0, 硬件环境为:CPU: Intel Xeon Gold 5218, 内存:64 GB, 显卡:NVIDIA GeForce RTX 2080Ti。

4.2 对比方法及参数设置

选取以下基线方法进行对比:BiGRU^[16]、BiGRU+Attention^[19]、TextCNN^[20]、ALBERT^[2]、ELECTRA^[21]、CAT-BiGRU^[22]、Cov-Att-Bilstm^[23]、CNN+BiGRU^[24]、SKEP^[9]、LSTM-GRU^[25]。若无特殊说明,所有方法默认迭代轮次 epochs=110, 批大小 batch_size=1 024, 学习率 learning_rate=0.000 2, L_2 惩罚系数 $\lambda=0.000 2$, 优化器为 Adam, 词向量通过 skip-gram 模型获得, 句子最大长度 max_len=100, 词嵌入维度 $m=95$, 属性向量维数 $n=5$ 。上述方法涉及的其他参数如表 3 所示。

表 3 不同方法的主要参数设置

Table 3 Main parameters for different models

分类方法	准确率
BiGRU ^[16]	BiGRU 层数 num_bigru=1
BiGRU+Attention ^[19]	BiGRU 维度 dim_bigru=100
TextCNN ^[20]	卷积核尺寸 filter_size=3,4,5 卷积核数量 num_kernel=100
ALBERT ^[2]	albert_small_zh_google 模型
ELECTRA ^[21]	electra_180g_small 模型
SKEP ^[9]	ernie_1.0_skep_large_ch 模型
CAT-BiGRU ^[22]	词向量维数 word2vec_size=200 句子最大长度 max_len=20
Cov-Att-Bilstm ^[23]	BiLSTM 层数 num_bilstm=2
CNN+BiGRU ^[24]	卷积核尺寸 filter_size=2,3,4,5 卷积核数量 num_kernel=100
LSTM-GRU ^[25]	词向量维数 word2vec_size=200
MFMS	温度系数 $T=3$, 权重因子 $a=0.7$

4.3 数据增强验证和测试集选定

为解决标记数据量不足的问题,利用 RoFormer-Sim 模型对训练文本进行数据增强,将训练数据分别扩增 1、2、3 倍。同时,根据评论发表时间选定测试集,将发表时间段为:2021.10.29 至 2021.11.04(最近 1 周)、2021.10.22 至 2021.11.04(最近 2 周)、2021.10.15 至 2021.11.04(最近 3 周)、2021.10.08 至 2021.11.04(最近 4 周)的标注评论数据分别作为测试集,即时间间隔 $D=1$ 周,时间阈值为最近 w 周($w=1,2,3,4$),对应样本数量分别为 407、801、1 074 及 1 309。

图 2 展示了 MFMS 方法在训练集数据增强 z ($z=0,1,2,3$) 倍时、针对不同时间区间(最近 w

周)测试集的情感分类结果。由图 2 可知,在测试数据集一致时,扩增的倍数在 1、2 倍时,情感分类性能有明显提升,说明 RoFormer-sim 生成模型能够一定程度地解决原本质量合格的标注评论较少,专有训练语料不足的问题,较好地提升情感分类性能。当扩增倍数为 $z=3$ 时,仅测试集为最近 2 周($w=2$)时,测试结果略有提升,在准确率和 F_1 值上分别仅提高了 0.07% 和 0.18%,其余测试集均有不同程度的下跌,这是因为较大的扩增倍数将引入较多的噪声样本,继而影响了情感分类的准确性。

进一步发现,当选用最近 1 周($w=1$)发表的评论为测试集,且训练集扩增倍数为 $z=2$ 时,其准确率和 F_1 值最高,分别为 86.6%、90.8%,比训练集为原数据($z=0$)时分别提高了 2.92%、2.06%。图 3 展示了 MFMS 方法在不同的文本增强倍数下、针对不同测试集时的平均训练时间,由图 3 可知,随着文本增强倍数的增加,模型的训

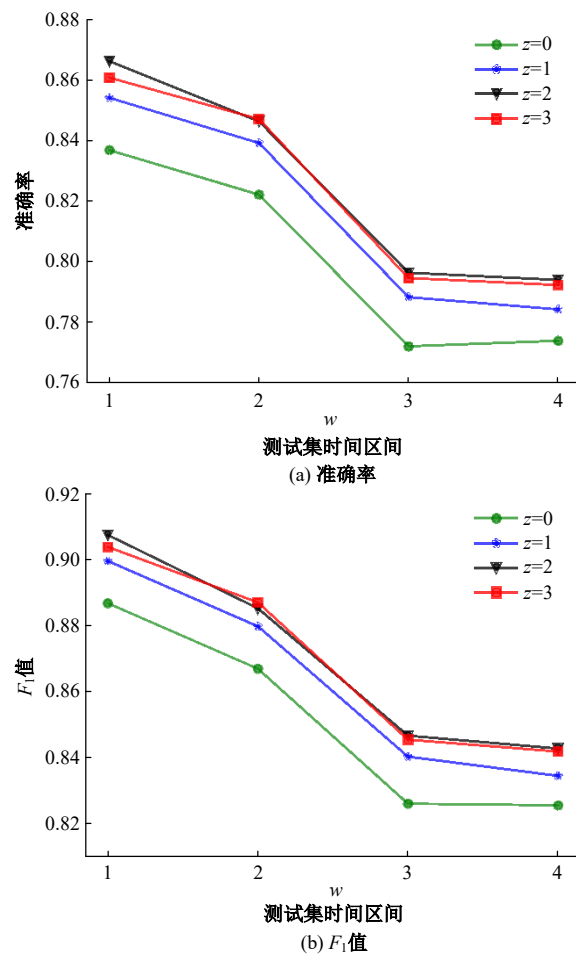


图 2 数据增强倍数及测试集对分类表现的影响

Fig. 2 Effect of data augmentation times and testing set on classification performance

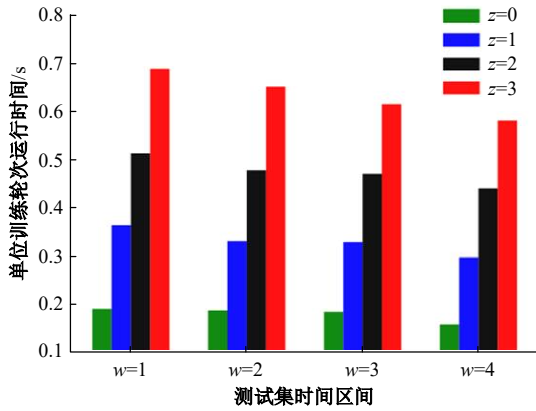


图 3 数据增强倍数及测试集对运行时间的影响

Fig. 3 Effect of data augmentation times and testing set on running time

练时间呈线性增长。因此,均衡考虑训练时间和分类效果,文本增强倍数、测试集评论的时间区间分别取 2 倍、最近 1 周,并将其应用于后续对比实验中。

4.4 epoch 影响分析

图 4 显示了训练轮数为 110 时,模型训练过程中的总损失、蒸馏损失、交叉熵损失、准确率和 F_1 值随训练轮次的变化情况。由图 4 可知,随着训练轮次的增加,各损失值均呈下降趋势,准确率和 F_1 值呈上升趋势。当 epoch 取值大于 20 时,准确率和 F_1 值趋于稳定,而损失值下降变得缓慢,说明此时模型的训练过程接近收敛状态。在各损失值中,总损失和交叉熵损失相近,且明显大于蒸馏损失,这是因为,由式(24)可知,总损失为蒸馏损失乘以 T^2 的结果、交叉熵损失以及惩罚项的加权和,使温度系数 T 限制了蒸馏损失的大小。

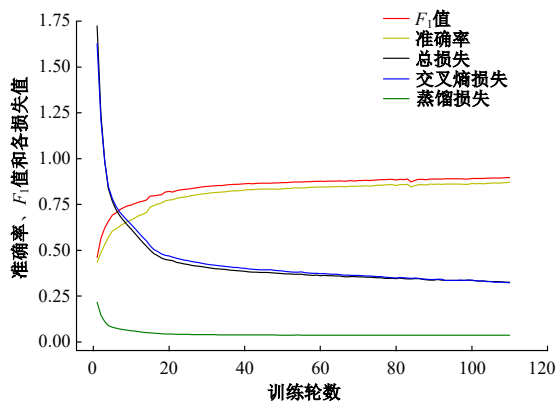


图 4 训练中各损失值、准确率和 F_1 值的变化趋势

Fig. 4 Trend of each loss value, accuracy value and F_1 value in training process

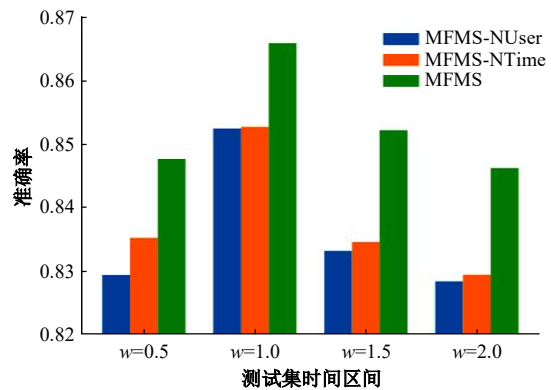
4.5 带时间属性的用户个性特征有效性验证

为了验证评论时间属性对提升情感分类性能的有效性,通过改变多特征模型的输入特征以获得不同的情感分类方法,具体如下:

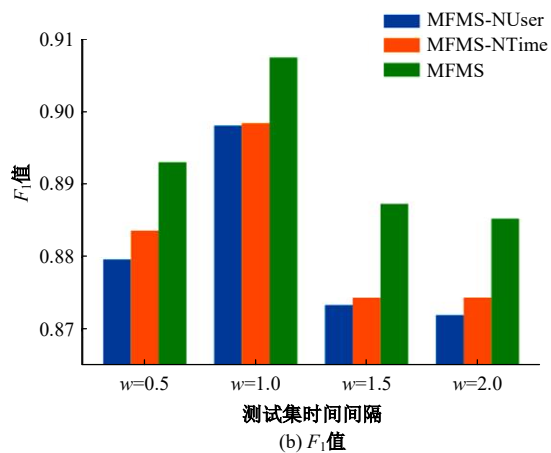
(1) MFMS-NTime: 与 MFMS 不同的是,该方法的 用户特征向量生成时,不考虑历史评论的发表时间因素,取历史评论向量的平均值得到。

(2) MFMS-NUser: 与 MFMS 不同的是,该方法不融合用户特征向量,仅融合用户当前评论向量和用户关系向量。

图 5 展示了各分类方法在测试集发表时间阈值为最近 w 周 ($w=0.5, 1, 1.5, 2$), 时间间隔 $D=0.5$ 周时的情感分类结果。由图 5 可知, MFMS 表现均优于 MFMS-NTime。当 w 值分别为 0.5、1、1.5、2 时, MFMS 相对于 MFMS-NTime 在 F_1 值上分别提高了 1.82%、1.35%、1.9%、1.77%, 验证了相比于不考虑历史评论的发表时间得到的用户特征向量,使用时间遗忘函数得到的用户特征向量对提升情感分类性能的有效性。当 w 值



(a) 准确率



(b) F_1 值

图 5 带时间属性的用户个性特征有效性验证

Fig. 5 Validation of user personality characteristics with time attribute

分别为 0.5、1、1.5、2 时, MFMS 相对于 MFMS-NUser 在准确率上分别提高了 0.93%、0.90%、1.30%、1.10%, 验证了融合基于用户历史评论和属性的用户特征向量对提升评论文本情感分类性能的有效性。进一步发现, 当测试集的时间间隔 w 值取 1 时, MFMS、MFMS-NTime、MFMS-NUser 在准确率与 F_1 值上均比 w 取其他值(0.5、1.5、2) 时有不同程度的提高, 进一步验证了本文把测试集时间区间选定为最近 1 周的合理性。

4.6 综合比较

表 4 列出了本文方法与经典深度学习方法在准确率、 F_1 值和 RMSE 值上的实验结果。由表 4 可知, 本文方法表现均优于预训练模型算法(ALBERT、ELECTRA 与 SKEP), 比预训练模型中表现最好的 ELECTRA 模型在准确率和 F_1 值上分别提升了 0.9% 和 1.0%, 在 RMSE 值上降低了 0.013。这是因为预训练模型仅将用户评论的语义信息作为特征, 忽略了用户的个性特征和社会关系特征, 因此, 降低了其在情感分类上的表现。

进一步分析表 4 可知, 在除预训练模型之外的深度学习方法中, CNN+BiGRU 模型表现最优, 而本文未经知识蒸馏的 MFMS 方法相比于 CNN+BiGRU 模型在准确率和 F_1 值上分别提升了 1.4% 和 1.0%, 在 RMSE 值上降低了 0.032。验证了在用户当前评论文本特征的基础上综合考虑带有时间属性的用户历史评论、属性特征和社交关系对提升情感分类性能的有效性。同时, 本文的 MFMS 方法比 MFMS 方法在准确率和 F_1 值

表 4 与典型深度学习方法的准确率、 F_1 和 RMSE 值比较
Table 4 Comparison of accuracy, F_1 and RMSE values of typical deep learning methods

分类方法	准确率	F_1	RMSE
BiGRU	0.830	0.881	0.412
BiGRU+Attention	0.834	0.882	0.407
TextCNN	0.826	0.882	0.418
ALBERT	0.841	0.885	0.397
ELECTRA	0.857	0.898	0.379
SKEP	0.835	0.881	0.406
CAT-BiGRU	0.812	0.868	0.433
Cov-Att-BiLSTM	0.824	0.878	0.419
CNN+BiGRU	0.841	0.890	0.398
LSTM-GRU	0.799	0.856	0.449
MFMS	0.855	0.900	0.380
MFMS	0.866	0.908	0.366

上分别提升 1.1% 和 0.8%, 比 SKEP 在准确率和 F_1 值上分别提升 3.1% 和 2.7%, 可见, 知识蒸馏理论使基于多特征融合的评论文本情感得分预测模型可以充分学习 SKEP 情感模型的通用性知识, 继而有效提高模型的分类准确率和泛化性。

4.7 模型通用性验证

为验证本文方法针对训练集外冷启动用户的情感分类性能, 通过修改本文方法以获得不同的情感分类方法。具体方法如下:

(1) MFMS-DW: 与 MFMS 不同的是, 该方法的用户关系向量采用 DeepWalk 算法生成。

(2) MFMS-NRel: 与 MFMS 不同的是, 该方法不融合用户关系向量, 仅融合用户当前评论向量和用户特征向量。

(3) MFMS-Attr: 与 MFMS 不同的是, 该方法通过求该冷启动用户与其他用户属性向量的余弦相似度得到最相似的前 5 个用户, 再取它们关系向量的平均值作为该冷启动用户的关系向量。

表 5 列出了不同方法对应的准确率和 F_1 值, 由表 5 可知, MFMS 表现优于 MFMS-NRel, 在准确率与 F_1 值上分别提高了 0.5% 与 0.4%, 验证了用户社会关系特征对提升情感分类性能具有较好的效果。MFMS 表现优于 MFMS-DW, 在准确率与 F_1 值上分别提高了 0.3% 与 0.2%, 这说明相比于 DeepWalk 算法, 使用 node2vec 算法得到的用户社会关系向量在提升情感分类性能上更加有效。进一步发现, MFMS 比未经知识蒸馏的 MFMS 方法在准确率与 F_1 值上分别提高了 1.3% 与 0.9%, 比教师模型 SKEP 在准确率和 F_1 值上分别提高了 2.8% 和 2.4%, 这主要是因为 MFMS 模型通过蒸馏训练学习到了 SKEP 模型中蕴含的通用性情感分类知识, 提高了模型对跨领域评论的检测能力。

为验证本文方法在其他数据集上的表现, 进

表 5 不同方法对应的准确率、 F_1 值

Table 5 Comparison of accuracy and F_1 values of different methods

分类方法	准确率	F_1
SKEP	0.835	0.881
MFMS-NRel	0.858	0.901
MFMS-Attr	0.862	0.903
MFMS-DW	0.860	0.903
MFMS	0.850	0.896
MFMS	0.863	0.905

一步选择英文 SST-2 数据集^[9]进行实验。选取其中 1/3 样本微调 SKEP 模型,剩余 2/3 样本作为本文方法数据集,在此基础上得到不同方法对应的准确率及 F_1 值,如图 6 所示。由图 6 可知,使用蒸馏模型的 MFMS 方法比未使用蒸馏模型的 MFM 方法在准确率和 F_1 值上分别提高了 1.2% 和 1.1%,这说明 MFMS 方法学习了 SKEP 模型蕴含的有用分类知识,有效提升了本文方法的情感分类效果。

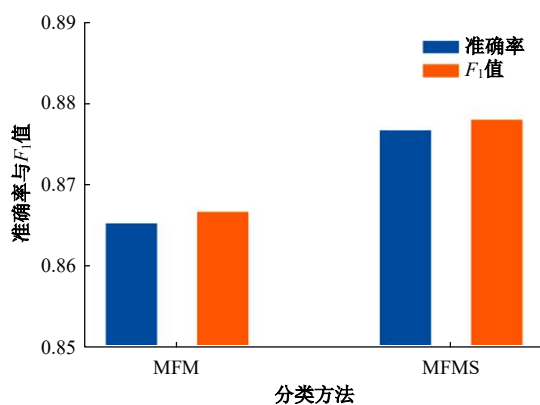


图 6 不同方法在 SST-2 上的准确率、 F_1 值

Fig. 6 Accuracy and F_1 values of different methods on SST-2

5 结束语

本文提出了一种基于知识蒸馏和评论时间的文本情感分类新方法,主要贡献包括:①利用 RoFormer-Sim 生成模型对标注评论进行数据增强,解决了训练数据不足的问题;②引入时间因素,为每个用户构建了具有遗忘属性的用户特征,并将其与用户社交关系特征、当前评论特征通过注意力机制融合,充分考虑不同类型特征对情感分类结果的影响;③提出了基于用户相似度的用户关系表示方法,提高了针对冷启动用户的表达能力,同时引入知识蒸馏的理论和训练方法,提高了模型的通用性和泛化性。在从股吧爬取的真实数据集上的实验结果表明:与传统典型方法相比,该方法在准确率及 F_1 值上获得明显提升。但是,本文方法未能有效考虑用户社会关系动态变化的影响,因此,未来将结合时序数据分析方法获得符合当前用户情感状态的社会关系特征,以进一步提高模型情感分类效果。

参考文献:

- [1] Umer M, Sadiq S, Nappi M, et al. ETCNN: extra tree and convolutional neural network-based ensemble model for COVID-19 tweets sentiment classification[J]. Pattern Recognition Letters, 2022, 164: 224-231.
- [2] Lan Z, Chen M, Goodman S, et al. ALBERT: a lite bert for self-supervised learning of language representations[C]//International Conference on Learning Representations, Addis Ababa, Ethiopia, 2020: 1-17.
- [3] Liu Z, Huang D, Huang K, et al. Finbert: a pre-trained financial language representation model for financial text mining[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, Yokohama, Japan, 2021: 4513-4519.
- [4] 赵亚欧, 张家重, 李贻斌, 等. 基于 ELMo 和 Transformer 混合模型的情感分析[J]. 中文信息学报, 2021, 35(3): 115-124.
Zhao Ya-ou, Zhang Jia-chong, Li Yi-bin, et al. Sentiment analysis based on hybrid model of elmo and transformer[J]. Journal of Chinese Information Processing, 2021, 35(3): 115-124.
- [5] Yang J, Zou X, Zhang W, et al. Microblog sentiment analysis via embedding social contexts into an attentive LSTM[J]. Engineering Applications of Artificial Intelligence, 2021, 97: 104048.
- [6] 蒋宗礼, 张静. 融合用户和产品信息的多头注意力情感分类模型[J]. 计算机系统应用, 2020, 29(7): 131-138.
Jiang Zong-li, Zhang Jing. Multi-head attention model with user and product information for sentiment classification[J]. Computer Systems & Applications, 2020, 29(7): 131-138.
- [7] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38-39.
- [8] 邵仁荣, 刘宇昂, 张伟, 等. 深度学习中知识蒸馏研究综述[J]. 计算机学报, 2022, 45(8): 1638-1673.
Shao Ren-rong, Liu Yu-ang, Zhang Wei, et al. A survey of knowledge distillation in deep learning[J]. Chinese Journal of Computers, 2022, 45(8): 1638-1673.
- [9] Tian H, Gao C, Xiao X, et al. SKEP: sentiment knowledge enhanced pre-training for sentiment analysis[J/OL]. [2023-07-25]. arXiv Preprint arXiv: 2005. 05635v2.
- [10] Turney P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[J]. Association for Computational Linguistics,

- 2002, 6: 417-424.
- [11] Xie W. Entity linking based on roformer-sim for chinese short texts[J]. *Frontiers in Computing and Intelligent Systems*, 2023, 4(1): 46-50.
- [12] Zhao Y, Liu S, Zhang Q, et al. Test case classification via few-shot learning[J]. *Information and Software Technology*, 2023, 160:107228.
- [13] Lewis M, Liu Y, Goyal N, et al. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension [C]//Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020: 7871-7880.
- [14] Ebbinghaus H. Memory: a contribution to experimental psychology[J]. *Annals of Neurosciences*, 2013, 20(4): 2004155.
- [15] Grover A, Leskovec J. Node2vec: scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2016: 855-864.
- [16] 乔百友, 武彤, 杨璐, 等. 一种基于BiGRU和胶囊网络的文本情感分析方法[J]. *吉林大学学报: 工学版*, 2024, 54(7): 2026-2037.
Qiao Bai-you, Wu Tong, Yang Lu, et al. A text sentiment analysis method based on BiGRU and capsule network[J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2024, 54(7): 2026-2037.
- [17] 王友卫, 童爽, 凤丽洲, 等. 基于图卷积网络的归纳式微博谣言检测新方法[J]. *浙江大学学报: 工学版*, 2022, 56(5): 956-996.
Wang You-wei, Tong Shuang, Feng Li-zhou, et al. New inductive microblog rumor detection method based on graph convolutional network[J]. *Journal of Zhejiang University (Engineering Science)*, 2022, 56(5): 956-996.
- [18] 陈洁, 王思雨, 赵姝, 等. 基于多粒度用户偏好的文档级情感分析[J]. *中文信息学报*, 2023, 37(7): 122-130.
Chen Jie, Wang Si-yu, Zhao Shu, et al. Multi-granular user preferences for document-level sentiment analysis[J]. *Journal of Chinese Information Processing*, 2023, 37(7): 122-130.
- [19] Li Y, Ni P, Li G, et al. Inter-personal relation extraction model based on bidirectional GRU and attention mechanism[C]//IEEE 5th International Conference on Computer and Communications (ICCC), Harbin, China, 2019: 1867-1871.
- [20] Guo B, Zhang C, Liu J, et al. Improving text classification with weighted word embeddings via a multi-channel textcnn model[J]. *Neurocomputing*, 2019, 363: 366-374.
- [21] Clark K, Luong M T, Le Q V, et al. Electra: pre-training text encoders as discriminators rather than generators[J/OL]. [2023-07-26]. arXiv Preprint arXiv: 2003. 10555.
- [22] Kamal A, Abulaish M. Cat-bigru: convolution and attention with bi-directional gated recurrent unit for self-deprecating sarcasm detection[J]. *Cognitive Computation*, 2022, 14: 91-109.
- [23] Ahmad W, Wang B, Martin P, et al. Enhanced sentiment analysis regarding COVID-19 news from global channels[J]. *Journal of Computational Social Science*, 2023, 6: 19-57.
- [24] Gao Z, Li Z, Luo J, et al. Short text aspect-based sentiment analysis based on CNN + BiGRU[J]. *Applied Sciences*, 2022, 12(5): 12052707.
- [25] Aslam N, Rustam F, Lee E, et al. Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model[J]. *IEEE Access*, 2022, 10: 39313-39324.