

情感先验的基层医疗服务评价风险挖掘方法

李至立¹ 付竹霓² 刘兴惠¹

LI Zhili FU Zhuni LIU Xinghui

摘要

基层医疗机构的舆情评价信息能够反映出医疗机构的服务质量,为挖掘出影响基层医疗机构服务质量的风险主题因素,文章提出一种情感先验的基层医疗服务评价风险挖掘方法。首先,利用微调的预训练 BERT 模型对评价文本进行特征提取,得到评价文本情感特征集;然后,利用支持向量机对样本进行情感分类,包括积极、中性和消极,将情感分类结果作为情感先验知识,得到包含中性和消极情感样本的情感先验语料库;最后,引入肘部法则进行自适应主题数量选定,利用隐含狄利克雷分布进行主题建模,最终得到风险主题。文章爬取大量基层医疗机构评价信息作为实验数据,大量实验结果表明,所提方法能够有效地挖掘出评价信息中的风险主题。

关键词

基层医疗;服务质量评价;LDA 主题挖掘;文本情感识别

doi: 10.3969/j.issn.1672-9528.2024.11.015

0 引言

随着全国基层医疗机构建设及技术水平的不断提升,基层医疗机构走向“医疗机构-患者-政府监督”的发展建设模式,同时也对医疗服务卫生建设有着直接影响^[1-2]。《国务院办公厅关于促进“互联网+医疗健康”发展的意见》^[3]《关于进一步完善医疗卫生服务体系的意见》^[4]等政策文件,指出了对全国医疗机构发展的指导意见,揭示了国家对医疗服务质量的高度重视。随着大数据、人工智能等新兴技术的不断发展,基层医疗信息化建设水平也不断提升,同时基层医疗机构服务质量监管也面临着新的问题和挑战。

文本主题挖掘作为自然语言处理领域的一个重要研究分支,能够从大量的文本数据中识别出相关主题,被广泛应用于网络舆情监测、情报分析、信息检索等^[5]。近几年,为了能够从文本信息中挖掘出定向主题信息,许多研究者将文本情感分析与主题挖掘技术相结合,将情感分类结果作为先验知识,挖掘出定向主题信息^[6],如微博舆情^[7-8]、App 用户反馈^[9]、在线教育评价^[10]等,本文将基于文本情感分析的主题挖掘技术定义为基于情感先验的主题挖掘(topic mining based on emotional priori, TMEP)方法。一些研究工作引入 TMEP 方法策略用以解决医疗服务质量监测问题,例如,曲丽萍等人^[11]通过建立情感语义分类词典,引入情感先验知识,利用相似度匹配方法挖掘出包含

情感语义的在线医疗服务评价主题;赵冬等人^[12]利用长短时记忆网络(long short-term memory, LSTM)对患者体验反馈文本进行情感分类,对于负面情感样本利用隐含狄利克雷分布(latent dirichlet allocation, LDA)进行主题建模,以挖掘医院服务问题;叶艳等人^[13]基于 LDA 和双向 LSTM 学习医疗服务质量主题的情感倾向分布,进一步对负面评论的服务质量主题进行筛选,以此挖掘出医疗服务质量问题。

现有的大多数 TMEP 相关方法通过预训练模型直接提取情感特征进行分类,或训练轻量级模型进行情感分类作为情感先验知识,这些做法会导致 TMEP 在特定场景数据下的情感分类结果效果不佳或模型泛化能力弱等问题。此外,已有的大多数 TMEP 方法在进行主题挖掘时直接利用 LDA 进行主题建模,需要进一步对 LDA 的主题数量进行实验分析,难以满足自适应主题挖掘的效果。

针对已有医疗服务质量评价的 TMEP 方法存在的问题,本文提出一种情感先验的基层医疗服务评价风险挖掘方法(risk mining method of service evaluation in primary medical institutions based on emotional priori, EmoSE-RM)。在 EmoSE-RM 方法中,首先,基于微调的预训练 BERT 模型^[14]提取情感特征,利用支持向量机(support vector machine, SVM)进行情感分类,将情感标签为消极和中立的样本作为情感先验语料库;然后在风险主题挖掘阶段,引入肘部法则^[15]进行风险主题数量的自主学习,构建自适应主题数 LDA(k-LDA)模型进行风险主题挖掘;最终

1. 山东纬横数据科技有限公司 山东烟台 264000

2. 烟台市芝罘区疾控中心 山东烟台 264000

挖掘到医疗服务质量评价中的风险主题。

1 情感先验的基层医疗服务评价风险挖掘方法

本文提出的 EmoSE-RM 方法框架如图 1 所示,主要分为情感先验语料库构建阶段和风险主题挖掘阶段。情感先验语料库构建阶段的主要任务是利用微调后的预训练 BERT 模型 (fine-tuned pre-trained BERT, FT-pre-BERT) 进行情感特征提取,然后利用 SVM 进行情感分类,将消极和中立情感的样本整理作为情感先验语料库;风险主题挖掘阶段的主要任务是改进传统的 LDA 主题建模过程,引入肘部法则自适应学习最优主题数,提出 k-LDA 主题建模方法,挖掘出服务质量风险主题。

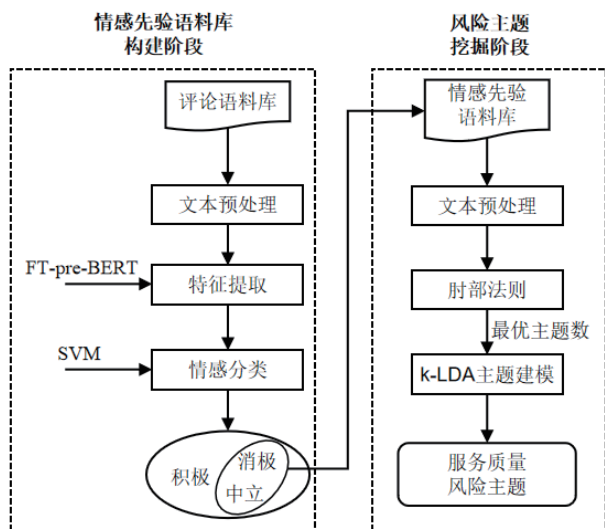


图 1 EmoSE-RM 方法框架图

1.1 情感先验语料库构建

1.1.1 微调预训练 BERT 模型

在搭建 EmoSE-RM 方法框架之前,首先需要利用情感标注的医疗服务质量评论数据对预训练的 BERT 模型进行微调,得到的 FT-pre-BERT 模型用于后续对评论文本数据进行情感特征提取,详细过程描述如下。

(1) 数据爬取与情感标注

为学习到更优的情感先验知识,本文爬取并筛选出 2892 条基层医疗机构服务质量评论数据,其中 1446 条进行人工情感标注,情感标签包括积极、消极和中立,用于微调 BERT,剩余的未被标注的 1446 条数据用于本文 EmoSE-RM 方法实验。

(2) 文本预处理

本文使用 Hugging 的 BertTokenizer 作为分词器,将文本分割成单词或字词,称之为 token,然后对 token 进行编码,转换为 BERT 模型可接受的输入格式,并为每个 token 分配唯一的标识符,用以之后的微调任务。

(3) 微调 BERT

本文使用 2018 年 Google 发布的预训练 BERT 模型 bert-base-uncased^[14] 进行微调,过程如图 2 所示。整个微调模型过程:首先输入预处理后的样本数据以及情感标签,经过 BERT 之后,得到能够表示词间语义及关系的 embedding,然后添加 softmax 层作为输出层,完成情感分类任务,最终得到的分类损失 loss 通过反向传播对模型参数进行更新,使模型参数适应 EmoSE-RM 方法中的情感分类任务。

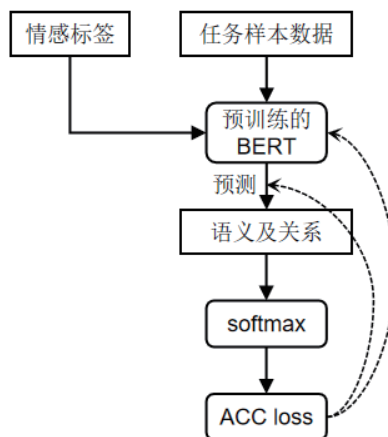


图 2 BERT 模型微调过程

1.1.2 SVM 情感分类

FT-pre-BERT 模型提取的情感特征共 768 维,在进行情感分类任务之前,需要进行特征降维。本文利用经典的无监督子空间学习方法,主成分分析 (principal component analysis, PCA)^[16],对高维的情感特征进行特征降维,去除冗余信息,节省计算资源。本文使用 PCA 保留 98% 的主成分,共得到 420 维特征。经过 PCA 降维后的情感特征作为评论情感特征集,记为 $X \in \mathbf{R}^{n \times d}$,其中, n 为样本数, d 为特征维度,用于 SVM 的训练和情感分类,计算公式为:

$$f(x) = \text{sgn}(W^T \phi(x) + b) = \text{sgn}\left(\sum_{i=1}^n a_i y_i K(x_i, x) + b\right) \quad (1)$$

式中: a_i 为拉格朗日乘子,样本 $x_i \in X$, 对应标签为 y_i , b 为偏置, $\phi(\cdot)$ 表示核映射, $K(\cdot)$ 代表核函数,本文使用线性核函数^[17]将样本的输入空间映射至高维空间,以解决多分类问题。

经过 SVM 分类后的样本情感分为积极、消极以及中立,由于潜在服务质量风险信息与消极、中立情感的样本有很大的关联,因此,本文筛选出消极、中立情感的样本,作为用于后续风险主题挖掘的语料库。由于该语料库经由情感知识先验得到,本文将该语料库称为情感先验语料库。

1.2 风险主题挖掘

1.2.1 自适应主题数学习

在风险主题挖掘阶段,以往的主题建模方法在主题建模后,需要进行实验并可视化分析以确定最优的主题数量,

本文对传统的 LDA 主题建模过程进行改进, 引入肘部法则^[15]和 K-means 聚类方法, 并设计最优主题数量计算函数, 自适应地选定 LDA 主题建模的最优数量, 详细过程描述如下。

(1) K-means 聚类

首先, 使用预训练的 BERT 模型 bert-base-uncased 对情感先验语料库样本进行特征提取, 得到特征集 $X^e \in \mathbf{R}^{n^e \times d}$, 其中, n^e 为情感先验语料库的样本数量, d 为特征维度。然后, 利用 K-means 算法样本划分为 k 个簇, 然后计算每个簇的质心, 记为 μ_j 。

使用簇内平方和 (within-cluster sum of squares, WCSS) 评价 K-means 聚类效果, 计算公式为:

$$WCSS(k) = \sum \sum (x_i^e - \mu_j)^2 \quad (2)$$

式中: k 表示聚类的簇数, x_i^e 表示第 i 个样本。

(2) 计算最佳主题数

肘部法则的思想策略是根据不同主题数量损失的变化曲线, 确定“肘部点”, 即为最优主题数量。本文引入肘部法则的思想, 将 WCSS 损失作为计算主题数量的度量损失, 通过计算以下优化目标函数, 找到 WCSS 变化的“肘部点”。

$$\arg \min_k \frac{|WCSS(k) - WCSS(k+1)|}{|WCSS(1) - WCSS(k+1)|} \quad (3)$$

式中: $|WCSS(k) - WCSS(k+1)|$ 表示相邻主题数量 WCSS 的变化值, 变化值越缓越接近“肘部点”, $|WCSS(1) - WCSS(k+1)|$ 表示初始状态到 $k+1$ 当前状态的 WCSS 下降量。本文将优化公式 (3) 的最小阈值设定为 0.001。通过最小化上述公式, 能够找到满足“肘部点”的 k 值。

1.2.2 k-LDA 主题建模

LDA 作为一种常用的主题建模方法, 其主要目的是学习出文本 - 主题 Dirichlet 分布 θ 以及主题 - 词 Dirichlet 分布 ϕ , 然后根据 ϕ 输出 k 个主题。与传统的 LDA 有所不同, 通过式 (3), 能够确定最优主题数 k , 作为改进后 LDA 算法的输入参数, 称之为 k-LDA, 其概率模型如图 3 所示。

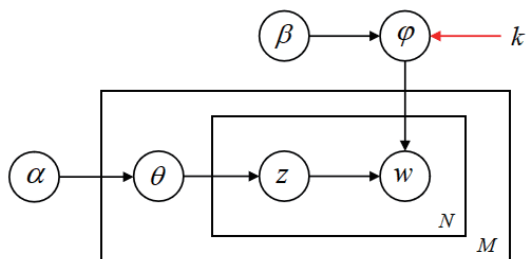


图 3 k-LDA 概率模型

其中, α 和 β 分别表示分布 θ 和分布 ϕ 的参数, z 表示

从分布 θ 采样出的主题, w 表示从分布 ϕ 采样出的词, M 代表文本数, N 代表某文本的关键词数。基于依赖关系构建联合概率分布为:

$$P(w, z, \theta, \phi | \alpha, \beta) = P(w | z, \phi) P(z | \theta) P(\theta | \alpha) P(\phi | \beta) \quad (4)$$

通过计算 k-LDA 的联合概率分布公式, 能够得到文本 - 主题概率分布 θ - 关键词概率分布 ϕ , 在给定主题数量 k 的前提下, 能够自适应地开展最优主题提取与分析。

2 实验

2.1 数据集与实验细节

本文爬取美团平台上部分基层医疗机构相关注册商户的评论信息, 包括美容美体店、按摩店、减肥店、口腔诊所、康复护理院等共计 8657 条数据, 经过筛选, 保留有效的评论信息, 共计 2892 条, 评价词云如图 4 所示。其中, 1446 条被标注的样本用于微调预训练 BERT 模型, 训练、测试 SVM 分类模型; 剩余的未被标注的 1446 条数据用于验证本文所提出的 EmoSE-RM 方法框架。



图 4 基层医疗机构服务质量评价词云

对于情感先验语料库构建阶段的 BERT 微调过程, 整个微调框架的环境为 PyTorch 1.11.0, 运行 GPU 算力为 NVIDIA V100-32GB, 使用 AdamW 优化器进行迭代优化, 前馈损失 loss 为分类损失 ACC, 取 80% 样本进行训练, 剩余 20% 用于测试, 参数经过调优设置为最佳, 学习率设置为 $1e-5$, epochs 设定为 50, 最终, 调优后的分类准确率为 96.51%。得到的 FT-pre-BERT 模型用于后续的评论文本情感特征提取。

对于 SVM 情感分类模型训练、测试以及风险主题挖掘, 均依赖 scikit-learn 1.5.1 环境, SVM 模型正则化参数设定为 1.0, 核函数 kernel='linear', K-means 方法以及隐含狄利克雷分布方法的随机采样种子 random_state=24, 最大迭代次数 max_iter=50。

2.2 实验结果及分析

2.2.1 情感分类及混淆矩阵

情感先验语料库的质量将在一定程度上影响风险主题挖掘的效果, 而 SVM 模型情感分类结果的优劣将直接影响情

感先验语料库的质量。本文使用 1446 条被标注的评论数据对 SVM 分类模型进行训练和测试实验。首先, 利用 FT-pre-BERT 模型提取 768 维情感特征, 然后利用 PCA 保留 98% 的能量, 将特征维度将至 420 维。使用分类准确率 ACC 作为实验结果的评价度量, 取 80% 样本用于训练, 剩余 20% 样本用于测试。

SVM 分类模型的混淆矩阵如图 5 所示, 横坐标为预测标签, 纵坐标为真实标签。首先, 积极情感的识别准确率为 96.51%, 消极情感的识别准确率为 96.36%, 中立情感的识别准确率为 97.84%; 其次, 消极情感的样本易与中立情感的样本混淆; 最后, SVM 对情感三分类的整体准确率为 96.90%。这些实验结果表明, SVM 模型能够很好的对评论文本进行情感的三分类, 用于本文所提方法的情感先验知识。

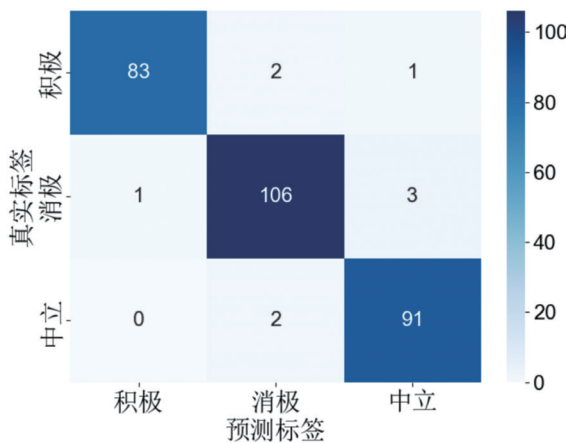


图 5 混淆矩阵

2.2.2 主题挖掘及分析

本文所提的 EmoSE-RM 方法框架中, 在情感语料库构建阶段, 使用 FT-pre-BERT 提取剩余的未被标注的 1446 条样本特征, 然后使用 PCA 进行特征降维, 再利用训练好的 SVM 模型进行情感分类, 最终, 积极情感的样本 406 条,

中立情感的样本 589 条, 消极情感的样本 451 条。将中立情感和消极情感的样本规整作为情感先验语料库, 共计 1040 条样本。

在风险主题挖掘阶段, 所提 EmoSE-RM 方法框架在本文所使用的情感先验语料库上的实验中, 学习到的自适应最优主题数量为 6, 各主题所覆盖的评论样本分布如图 6 所示, 本文所提算法学习到的基层医疗机构服务质量评价风险主题-关键词矩阵如表 1 所示。

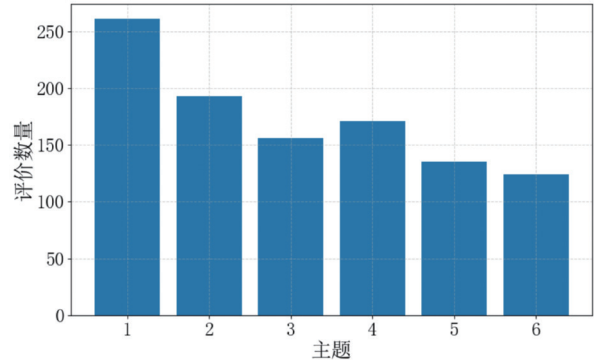


图 6 主题样本数量分布

总结分析图 5 及表 1 中的数据, 可以得出以下几条重要结论。

(1) 主题 1、主题 2 以及主题 4 的占比分别为 0.250 9、0.185 5、0.164 4, 总占比为 0.600 8, 能够覆盖 60% 的评论文本, 其中的风险关键词主要包括卫生、设施、服务、位置、服务态度、环境、体验、前台、服务员等, 从中可以看出, 现有的基层医疗机构在这些方面存在一定的风险问题, 会直接影响基层医疗机构的服务质量以及消费者的评价。

(2) 主题 3、主题 5、主题 6 的占比分别为 0.150 0、0.129 8、0.119 2, 总占比为 0.399, 能够覆盖 40% 的评论文本数据, 其中的风险关键词主要包括头发、技师、按摩、手法、技术、时间等, 从中可以看出, 这些方面存在的问题也将会在一定程度上影响基层医疗机构的服务质量。

表 1 基层医疗机构服务质量评价风险主题 - 关键词矩阵

主题 1 (0.250 9)		主题 2 (0.185 5)		主题 3 (0.150 0)		主题 4 (0.164 4)		主题 5 (0.129 8)		主题 6 (0.119 2)	
关键词	词频	关键词	词频	关键词	词频	关键词	词频	关键词	词频	关键词	词频
不好	0.139 8	体验	0.025 1	头发	0.044 7	服务	0.137 3	技师	0.032 9	没有	0.080 9
卫生	0.132 7	一个	0.021 8	没有	0.029 7	前台	0.069 5	服务	0.021 5	太差	0.041 5
设施	0.063 2	这个	0.021 1	还有	0.027 0	态度	0.066 2	按摩	0.019 7	设施	0.041 2
服务	0.048 0	没有	0.018 5	差评	0.026 8	服务态度	0.045 4	手法	0.016 8	差劲	0.036 2
位置	0.040 1	自己	0.017 7	体验	0.024 7	不好	0.032 0	技术	0.016 8	什么	0.021 6
服务态度	0.037 3	但是	0.017 2	太差	0.020 3	冷漠	0.026 3	时间	0.016 1	最差	0.019 2
环境	0.031 0	不是	0.016 1	效果	0.020 0	服务员	0.025 5	没有	0.016 0	就是	0.018 7
非常	0.017 4	还是	0.013 6	一般	0.017 1	差评	0.018 4	可以	0.015 8	热水	0.018 4
太差	0.015 1	知道	0.013 0	一次	0.017 0	特别	0.018 1	分钟	0.015 5	真的	0.015 4
干净	0.014 6	可以	0.012 8	特别	0.017 0	差劲	0.017 0	环境	0.015 4	环境	0.015 2

(3) 综合主题占比及关键词词频数据, 基层医疗机构以及监管部门, 应注意以下方面存在的风险问题, 风险依次降序, 包括卫生、服务、前台、态度、设施、服务态度、位置、技师、环境、体验、技术、时间。基层医疗机构可参考上述关键点, 按照优先次序进行长期自监督和改善管理。

3 结论

本文提出了一种情感先验的基层医疗服务评价风险挖掘方法 (risk mining method of service evaluation in primary medical institutions based on emotional priori, EmoSE-RM)。首先利用微调后的 BERT 模型提取未标注评论数据的情感特征, 然后使用训练好的 SVM 模型对文本数据进行情感分类, 将中立和消极情感的样本整合成情感先验语料库用于风险主题挖掘。然后对传统的 LDA 算法策略进行改进, 提出 k-LDA, 引入 K-means 与肘部法, 自适应学习文本最优主题数量 k , 开展自适应风险主题挖掘。最后大量的实验结果及结论表明, 本文所提的 EmoSE-RM 方法能够高效地挖掘出基层医疗机构服务质量风险主题, 给基层医疗机构及监管部门提供参考建议。在未来, 将探索更细粒度的情感分类方法, 学习更为丰富的情感先验知识, 并尝试收集整理更多的情感评论样本, 引入深度学习框架进行风险主题挖掘, 获得更好的效果。

参考文献:

- [1] 牛亚冬, 张研, 叶婷, 等. 我国基层医疗卫生机构医疗服务能力发展与现状 [J]. 中国医院管理, 2018, 38(6): 35-37+41.
- [2] 彭建军, 汪化睿. 公共服务均等化促进中华民族共同体建设的内在逻辑: 以医疗卫生为例 [J]. 中南民族大学学报 (人文社会科学版), 2024, 44(3): 28-37+182.
- [3] 国务院办公厅关于促进“互联网+医疗健康”发展的意见 [J]. 中华人民共和国国务院公报, 2018(14): 9-13.
- [4] 中共中央办公厅国务院办公厅印发《关于进一步完善医疗卫生服务体系的意见》[J]. 中华人民共和国国务院公报, 2023(10): 8-13.
- [5] 许海云, 董坤, 刘春江, 等. 文本主题识别关键技术研究综述 [J]. 情报科学, 2017, 35(1): 153-160.
- [6] 朱晓霞, 宋嘉欣, 张晓缙. 基于主题挖掘技术的文本情感分析综述 [J]. 情报理论与实践, 2019, 42(11): 156-163.
- [7] 王晰巍, 李玥琪, 刘婷艳, 等. 新冠肺炎疫情微博用户情感与主题挖掘的协同模型研究 [J]. 情报学报, 2021, 40(3): 223-233.
- [8] 曾莉, 杨添宝, 周慧. 基于 LDA 与注意力机制 BiLSTM 的

微博舆情分析模型 [J]. 南京理工大学学报, 2022, 46(6): 742-748.

- [9] 吴华君, 何聚厚, 陈其铁, 等. 面向职业教育在线精品课程评价的情感分析与主题挖掘 [J]. 中国职业技术教育, 2022(2): 55-63.
- [10] 沈棋琦, 张雪, 赵杨. 基于评论主题挖掘的移动图书馆 App 用户情感体验评价体系构建 [J]. 图书馆学研究, 2023(4): 41-51.
- [11] 由丽萍, 王世钰. 基于框架语义的在线医生服务评价主题识别 [J]. 情报理论与实践, 2019, 42(9): 166-170.
- [12] 赵冬, 马敬东, 罗玮, 等. 基于情感分类和主题建模的患者体验分析 [J]. 中国卫生信息管理杂志, 2019, 16(4): 502-507.
- [13] 叶艳, 吴鹏, 周知, 等. 基于 LDA-BiLSTM 模型的在线医疗服务质量识别研究 [J]. 情报理论与实践, 2022, 45(8): 178-183+168.
- [14] GEETHA M P, RENUKA D K. Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model [J]. International Journal of Intelligent Networks, 2021, 2(1): 64-69.
- [15] SHI C M, WEI B T, WEI S L, et al. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm [J]. EURASIP journal on wireless communications and networking, 2021, 2021: 31-47.
- [16] WOLD S, ESBENSEN K, GELADI P. Principal component analysis [J]. Chemometrics and intelligent laboratory systems, 1987, 2: 37-52.
- [17] 戴晓娟. 基于 SVM 线性核函数情感分类模型的建立和研究 [J]. 哈尔滨师范大学自然科学学报, 2014, 30(3): 55-57.

【作者简介】

李至立 (1988—), 男, 山东济宁人, 硕士, 高级工程师, 研究方向: 人工智能与大数据。

付竹霓 (1970—), 女, 山东烟台人, 硕士, 副主任技师, 研究方向: 生物化学。

刘兴惠 (1985—), 通信作者 (email: liuxh@vhengdata.com), 男, 山东滨州人, 硕士, 高级工程师, 研究方向: 人工智能与大数据。

(收稿日期: 2024-07-30)