

基于改进模糊神经网络的社交网络数据深度挖掘方法

马振尧¹ 郭少芳²
MA Zhenyao GUO Shaofang

摘要

社交网络数据高维且多样。在处理大规模数据时,未裁剪的模糊神经网络冗余多,增加复杂性,难捕关键特征,影响挖掘结果准确性。为此,文章提出了一种基于改进模糊神经网络的社交网络数据深度挖掘方法。通过网络爬虫从社交平台获取用户信息、交互记录及内容等原始数据,并进行清洗、离散化、归一化及模糊化处理,以降低噪声、冗余和复杂性,满足改进模糊神经网络的输入要求。构建并改进模糊神经网络,采用网络裁剪和引入注意力机制简化网络结构并增强对重要信息的捕捉能力,在训练中利用模糊化函数和反向传播机制,提取社交网络的潜在规律和特征。以提取的网络知识与规律特征为依据,对社交网络数据进行深度挖掘,得出反映社交网络深度特征的挖掘结果,并进行冗余检测和结果更新,以保证挖掘质量。测试实验得出:与传统挖掘方法相比,设计方法的 AUC 值和全局聚类系数取值更大,表明具有更优的挖掘精度和深度。

关键词

改进模糊神经网络; 社交网络; 网络数据; 网络爬虫; 深度挖掘

doi: 10.3969/j.issn.1672-9528.2025.11.038

0 引言

社交网络是现代社会重要的在线交流平台,允许用户通过互联网建立、维护和扩展个人及群体的联系。社交网络通常提供各种功能,如发布状态更新、分享照片和视频、参与讨论组、关注感兴趣的话题及人物等^[1]。为能够从大量的社交数据提取有价值的信息,以便为组织和个人提供有价值的洞察力,为了能够给商业决策、个性化推荐提供数据参考,提出社交网络数据深度挖掘方法。

社交网络数据深度挖掘是利用先进的数据挖掘、机器学习和统计分析技术,对社交网络中的大量数据进行深入分析和挖掘的过程。现阶段发展较为成熟的数据挖掘方法中,文献[1]提出的基于决策树的社交网络数据挖掘方法将社交网络视作多维向量空间,通过分析特定维度来界定用户的兴趣范围和兴趣焦点。在明确样本属性集合的基础上,依据已有行为信息构建测试路径,计算路径下各子集属性的重要性,并迭代此过程直至识别出具有相同属性的数据项。在实际应用中,社交网络数据中的用户行为具有动态性和模糊性,导致测试路径的构建不够准确,从而影响关键特征的提取,使得挖掘结果可能滞后或不准确。文献[2]提出的基于深度学习模型的数据挖掘方法构建并训练 BiLSTM-CRF 模型,在确

定 LDA 模型的最佳主题数情况下,采用困惑度与主题方差相结合的方法,构建 BiLSTM-CRF-LDA 模型,用以挖掘相关数据中的信息。但社交网络数据的多样性和模糊性可能导致 LDA 模型的主题划分不够准确,从而影响关键特征的提取,进而影响挖掘结果的实际应用价值。

针对上述不足,引入改进模糊神经网络以优化挖掘效果。模糊神经网络是模糊理论同神经网络相结合的产物,它汇集了神经网络与模糊理论的优点,集学习、联想、识别、信息处理于一体。通过模糊神经网络的改进与应用,优化设计社交网络数据深度挖掘方法,以期得出满足质量要求的数据挖掘结果。

1 社交网络数据深度挖掘方法设计

1.1 社交网络数据预处理与模糊化转换

社交网络数据富含用户行为、关系、文本、图像视频等信息,但易受噪声、缺失、异常值干扰,且维度高、冗余复杂,挖掘难度大。改进模糊神经网络需数据预处理与模糊化转换:预处理旨在清理数据、降噪、降维;模糊化则将不确定数据转化为模糊集合,保留模糊性信息。此过程确保数据质量,减少分析偏差,为后续分析奠定坚实基础。

在社交网络环境中,利用网络爬虫程序抓取网络,获取社交网络数据样本,初始获取的社交网络数据样本包括用户 ID、好友关系、帖子内容等,用公式表示为:

1. 韩国国立韩巴大学 大田广域 34158

2. 安徽理工大学 安徽淮南 232000

$$X = y_{\text{crawler}}(X_{\text{net}}) \quad (1)$$

式中: X_{net} 为社交网络中的实际存储数据; $y_{\text{crawler}}()$ 为网络爬虫函数。

为满足改进模糊神经网络的运行要求, 需对初始获取的社交网络数据样本进行预处理, 包括离散化处理和归一化处理两个步骤, 处理过程可量化为:

$$\begin{cases} x_{\text{discrete}} = \frac{\delta}{\bar{x}} \\ x_{\text{normalization}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \end{cases}, x \in X \quad (2)$$

式中: δ 和 $\bar{x}$ 分别表示社交网络数据的标准差和平均值; $x_{\max}$ 和 $x_{\min}$ 分别表示社交网络数据的最大值和最小值, 将数据预处理结果重新赋值给初始样本数据^[2]。

数据模糊化处理是满足改进模糊神经网络中的模糊规则, 处理结果为:

$$x_{\text{vague}} = y_s(y_{\text{vague}}(x)) \quad (3)$$

式中: $y_s()$ 为隶属度函数; $y_{\text{vague}}()$ 为模糊转换函数。

1.2 利用改进模糊神经网络深度提取网络知识与规律

1.2.1 模糊神经网络构建

在社交网络数据的深度挖掘中, 模糊神经网络的应用提供有效的手段处理数据的复杂性和非线性关系。通过将神经网络的基本组成单元(输入层、隐藏层、输出层)与模糊模块(输入、模糊化、模糊规则、解模糊)相结合, 构建适用于社交网络数据分析的模糊神经网络, 如图1所示。

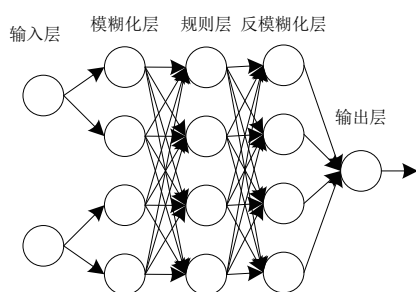


图1 模糊神经网络结构图

图1展示模糊神经网络的五个组成部分: 输入层接收原始数据; 模糊化层转换数据为模糊集合隶属度, 适应模糊逻辑; 规则层用预设模糊规则推理; 去模糊化层转模糊结果为清晰数值; 输出层给予最终训练结果, 如用户行为预测、社交趋势分析等。

1.2.2 模糊神经网络改进

在社交网络数据挖掘中, 初始模糊神经网络庞大复杂, 增加计算负担且规则难解。为此, 采用网络裁剪简化结构, 移除非关键神经元; 并引入注意力机制提升重要信息捕捉能力, 确保挖掘结果的准确性和可靠性^[3-5]。

模糊神经网络的裁剪包括两个步骤为去除多余的隐藏层神经元和削减不必要的神经元间连接。在实际的模糊神经网络裁剪中, 针对社交网络的数据特性, 设定两个随机正标量 β_1 和 β_2 , 设定的正标量满足条件:

$$\beta_1 + \beta_2 < 0.5 \quad (4)$$

提取初始构建模糊神经网络中各层权重值为 (ω, ω) , 这些权重值反映神经网络对社交网络数据特征的初始学习结果。模糊神经网络的裁剪条件为:

$$\begin{cases} \max |\omega_{ij}| \leq 4\beta_1 \\ |\omega_k| \leq 4\beta_2 \end{cases} \quad (5)$$

式中: ω_{ij} 和 ω_k 分别表示模糊神经网络中第 i 和第 j 层之间的连接权重以及神经元 k 的权重值。

若当前模糊神经网络层或节点满足式(5), 表示裁剪条件, 则对相应连接以及神经元进行删除处理, 否则无须执行删除操作^[6-7]。重复执行上述操作, 将删除对象的前后对象进行连接, 以确保神经网络的连通性和对社交网络数据特征的学习能力不受影响。为提升模糊神经网络在社交网络数据深度挖掘中的性能, 引入注意力机制, 实现对模糊神经网络中各神经元或组成层中权重值的调整, 增强模糊神经网络对重要信息的捕捉能力。

通过结合社交网络数据的特性进行模糊神经网络的裁剪, 并引入注意力机制进行优化, 能够构建高效和准确的网络, 为社交网络数据的深度挖掘提供有力支持。

1.2.3 训练改进模糊神经网络获取网络知识与规律特征

训练时处理后的社交网络数据输入改进模糊神经网络。模糊化层用高斯函数等转换数据以适应模糊逻辑。规则层结合注意力机制权重进行推理, 去模糊化层转换结果为清晰数值。这是模糊逻辑与神经网络结合的关键, 输出有意义结果。最终输出层提供用户行为预测、社交趋势分析、用户群体划分等特征提取结果, 通过反向传播调整参数, 直至完成迭代。模糊化层的训练过程为:

$$u_{\text{Fuz}} = h_{\text{Fuz}}(x_{\text{vague}}) = \exp \left\{ - \left(\frac{x_{\text{vague}} - c}{\frac{1}{\kappa_{\text{distribution}}}} \right)^2 \right\} \quad (6)$$

式中: $h_{\text{Fuz}}()$ 为模糊化函数; c 为高斯函数的中心值; $\kappa_{\text{distribution}}$ 为模糊化函数的尺度因子。

将模糊化层的输出结果输入到规则层中, 输出结果为:

$$u_{\text{rule}} = \mathfrak{g}_{\text{rule}} \cdot \exp \left\{ - \frac{(u_{\text{Fuz}} - \sigma)^2}{2\chi^2} \right\} \quad (7)$$

式中: σ 和 χ 分别为均值参数和标准差参数; $\mathfrak{g}_{\text{rule}}$ 为模糊化层中的权重; 结合注意力机制确定权重 $\mathfrak{g}_{\text{rule}}$ 的取值为:

$$g_{\text{rule}} = h_{\text{softmax}} \left(\frac{s(q, r)}{\sqrt{e}} \right) \quad (8)$$

式中: $h_{\text{softmax}}()$ 为 softmax 函数; $s(q, r)$ 为查询与键之间的相关性, q 和 r 分别为查询向量和键向量; e 为查询向量和键向量的维度^[8-10]。

将式(8)的计算结果代入到式(7)中,即可得出规则层的输出结果,将其输入到去模糊化层中,得出输出结果为:

$$u_{\text{Deblurring}} = \frac{\int u_{\text{rule}} h_{\text{Fuz}}(u_{\text{rule}}) du_{\text{rule}}}{\int h_{\text{Fuz}}(u_{\text{rule}}) du_{\text{rule}}} \quad (9)$$

将去模糊化层的运算结果输入到输出层中,最终输出的改进模糊神经网络训练学习输出结果即为社交网络知识与规律特征的提取结果,即:

$$\tau = u_{\text{Deblurring}} \quad (10)$$

在实际训练中,利用神经网络模块中的反向传播机制,对相关训练参数与权重进行调整,当训练次数达到最大迭代次数时,即得出社交网络知识与规律特征的最终获取结果。

1.3 社交网络数据深度挖掘实现

按照图2对社交网络数据进行挖掘。挖掘结果反映社交网络的深度特征。为保证挖掘质量,进行冗余检测并剔除冗余数据,根据社交网络动态更新结果,完成深度挖掘操作。

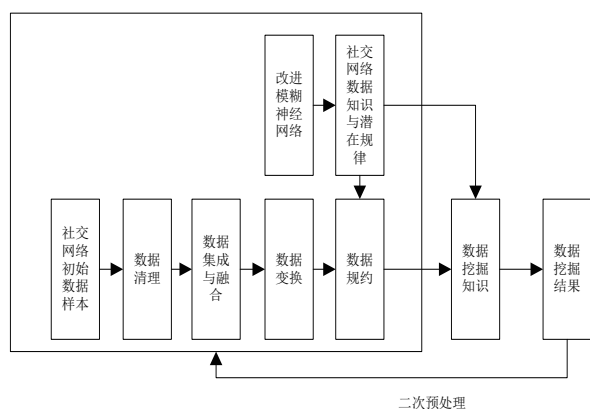


图2 社交网络数据深度挖掘流程图

按照图2流程,得出社交网络数据的深度挖掘结果为:

$$x_{\text{dig}} = \tau \cdot X^{\lambda} \quad (11)$$

式中: λ 为社交网络数据的挖掘深度。为保证挖掘质量,需对初始挖掘数据进行冗余检测,并剔除冗余数据。

2 挖掘性能测试实验分析

为测本文方法的性能设计测试实验,选择某社交网络实验,用其后台数据库为数据源。定挖掘目标,用设计方法得结果,与期望结果比对得性能。该网络有85万用户,62%月活,日发信息4.5万条。后台为MySQL,数据15.7GB。随机设挖掘任务,记结果与数量。为确保设计方法调用改进

模糊神经网络算法,对相关运行参数进行设定。设定模糊隶属函数为 Gaussian 函数,神经元总数量为100个,初始权重为0.2,学习率为0.01,迭代次数为500。改进模糊神经网络算法运行过程的训练情况,如图3所示。

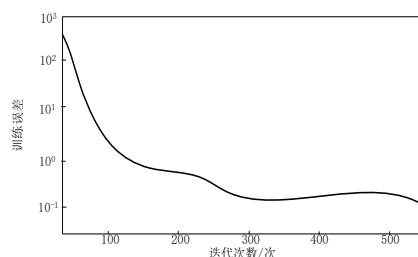


图3 改进模糊神经网络算法训练过程图

通过图3可知通过500次迭代训练,算法的训练误差降低,证明改进模糊神经网络算法能够高效处理社交网络数据,提取知识与规律特征,为社交网络分析和应用提供支持。

为满足改进模糊神经网络算法和社交网络数据深度挖掘方法的运行,对挖掘方法的硬件运行环境进行配置。优化设计的基于改进模糊神经网络的社交网络数据深度挖掘方法选用 Neuzz 工具作为开发测试工具,具有神经网络辅助的模糊测试能力,可启发类似技术在社交网络数据挖掘中的应用。将开发的数据深度挖掘应用到选择的社交网络及其数据库中,得出社交网络数据深度挖掘结果,如图4所示。

用户ID	用户名	注册时间	好友数量	发布信息数量
S0001	U123456	2023-01-15	56	11245
S0002	Alice	2021-11-30	89	5878
S0003	B527	2021-09-25	134	4736
S0004	Carol	2022-07-22	123	1905
S0005	#2567	2023-03-08	67	11200
S0006	Frank	2005-04-26	98	24598
S0007	Grace	2012-06-29	45	15479
S0008	Q4376	2015-03-12	26	9877

```

class AdaptiveFuzzyLayer(tf.keras.layers.Layer):
    def __init__(self, units):
        super(AdaptiveFuzzyLayer, self).__init__(units)
        self.units = units
    
```

图4 社交网络数据深度挖掘结果

按照社交网络数据挖掘任务的执行结果。设置传统的基于决策树的社交网络数据挖掘方法和基于深度学习模型的数据挖掘方法作为实验对比,在相同实验环境下完成对比方法的开发,并得出数据挖掘结果。

为实现基于改进模糊神经网络的社交网络数据深度挖掘方法挖掘性能的量化测试,设置 AUC 值作为数据挖掘精度的测试指标,测试结果为:

$$\alpha = \frac{1}{2} \sum_{i=1}^{n_{\text{ROC}}} (\eta_{\text{false}}(i+1) - \eta_{\text{false}}(i)) \cdot (\eta_{\text{real}}(i+1) - \eta_{\text{real}}(i)) \quad (12)$$

式中: n_{ROC} 为 ROC 曲线上的数据点数量; $\eta_{\text{false}}(i)$ 和 $\eta_{\text{real}}(i)$ 为第 i 个数据点的假正率和真正率。

在实际测试中,通过输出挖掘数据结果与期望数据结果

比对, 确定当前数据点的真、假属性, 由此绘制 ROC 曲线, 曲线在构成梯形的面积之和即为 AUC 值, 测试得出 AUC 值越高, 证明对应方法的挖掘精度越优。此外, 设定全局聚类系数作为数据挖掘深度的测试指标, 该指标的测试结果为:

$$\phi = \frac{\sum \phi(j)}{n_{\text{net}}} \quad (13)$$

式中: $\phi(j)$ 为社交网络挖掘数据节点 j 的局部聚类系数; n_{net} 为挖掘的网络数据节点数量。

较高的聚类系数意味着数据点之间的连接更为紧密, 反映社交网络中的社区结构或群体特征更加明显, 具有更优的数据挖掘深度。

通过数据统计, 得出反映数据挖掘精度性能的测试结果, 如图 5 所示。

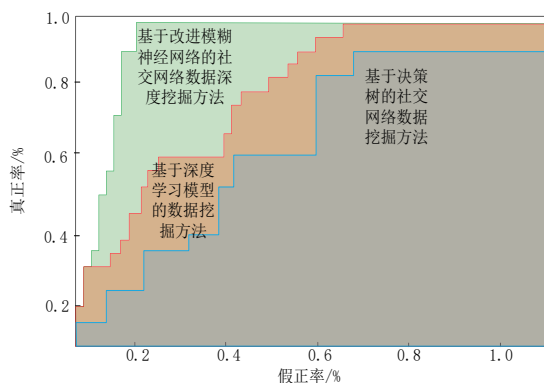


图 5 社交网络数据挖掘精度测试结果

图 5 看出与两种传统方法相比, 本文方法对应 ROC 曲线构成面积更大, 即 AUC 值取值更大。另外, 通过数据的统计与式 (13) 的计算, 得出数据挖掘深度的测试结果如表 1 所示。

表 1 社交网络数据挖掘深度测试数据表

数据挖掘任务编号	基于决策树的社交网络数据挖掘方法的全局聚类系数	基于深度学习模型的数据挖掘方法的全局聚类系数	基于改进模糊神经网络的社交网络数据深度挖掘方法的全局聚类系数
1	0.7	0.8	1.0
2	0.5	0.7	0.9
3	0.6	0.6	1.0
4	0.6	0.8	1.0
5	0.5	0.7	0.9

通过平均值计算 3 种挖掘方法全局聚类系数的平均值为 0.58、0.72 和 0.96。其中, 本文方法的全局聚类系数平均值为 0.96, 高于其他方法, 证明在社交网络数据深度挖掘中的优势。本文方法结合神经网络的自适应学习能力和模糊逻辑的表达能力, 能够处理社交网络数据, 提取潜在规律和特征。

3 结语

本文研究的改进模糊神经网络融合模糊逻辑与神经网络的优势, 增强了对社交网络数据中不确定性和模糊性的处理能力。通过引入网络裁剪和注意力机制, 进一步优化模型结构, 降低了冗余和噪声干扰, 从而实现对关键特征的精准捕捉和潜在规律的深度提取。该方法在挖掘精度和深度上均优于传统方法, 展现出更高的实用价值和广泛的应用前景。

参考文献:

- [1] 韩永印, 王侠, 王志晓. 基于决策树的社交网络隐式用户行为数据挖掘方法 [J]. 沈阳工业大学学报, 2024, 46(3): 312-317.
- [2] 肖琪耀, 贾宝山, 徐以诺, 等. 基于深度学习模型的煤矿安全隐患数据主题挖掘 [J]. 中国安全生产科学技术, 2024, 20(4): 49-55.
- [3] 易庚, 何琳, 刘锦明, 等. 基于迁移学习算法的电力数据挖掘模型 [J]. 沈阳工业大学学报, 2023, 45(5): 510-515.
- [4] 毛清华, 陈彦璋, 马骋, 等. 基于模糊神经网络 PID 的煤矿掘进机俯仰控制研究 [J]. 工矿自动化, 2024, 50(8): 135-143.
- [5] 王春艳, 金鹏, 桂琪皓. 区间二型模糊神经网络遥感图像分割方法 [J]. 测绘科学, 2024, 49(5): 84-98.
- [6] 张勇飞, 陈涛. 基于模糊神经网络的舰船雷达图像弱小目标检测 [J]. 舰船科学技术, 2024, 46(9): 147-150.
- [7] 夏小雅, 赵生字, 韩凡宇, 等. 面向开源协作数字生态的信息服务与数据挖掘 [J]. 计算机科学, 2024, 51(10): 187-195.
- [8] 周燕, 肖莉. 基于改进关联聚类算法的网络异常数据挖掘 [J]. 计算机工程与设计, 2023, 44(1): 108-115.
- [9] 马翔悦, 徐金东, 倪梦莹. 基于多尺度特征模糊卷积神经网络的遥感图像分割 [J]. 山东大学学报 (工学版), 2024, 54(3): 44-54.
- [10] 白仲航, 项钰, 谭昭芸, 等. 自适应神经模糊推理系统优化的快速上肢评估方法 [J]. 计算机集成制造系统, 2024, 30(5): 1643-1656.

【作者简介】

马振尧 (1997—), 男, 山东东营人, 博士研究生, 研究方向: 大数据分析。

郭少芳 (1990—), 女, 山西太原人, 博士, 讲师, 研究方向: 新媒体传播。

(收稿日期: 2025-03-10 修回日期: 2025-11-05)