

文章编号: 2095-2163(2022)12-0214-04

中图分类号: TP391

文献标志码: A

基于分块检测的社区网络敏感信息聚类算法

王学军

(广州华立学院, 广州 511325)

摘要: 为了提高无线虚拟社区网络敏感特征信息聚类能力,需要进行数据优化聚类处理,提出基于分块文本相似度检测的无线虚拟社区网络敏感特征信息网格强化聚类算法。采用异构有向图分析方法进行无线虚拟社区网络敏感特征信息存储结构设计,结合特征空间重组技术进行无线虚拟社区网络敏感特征信息结构重组,提取无线虚拟社区网络敏感特征信息的关联信息特征量,采用分块文本相似度检测的方法实现对社区网络敏感特征信息谱密度特征提取和融合聚类处理。仿真结果表明,采用该方法进行社区网络敏感特征信息谱密度融合的聚类性较好,对社区网络敏感信息的分块检测能力较强。

关键词: 社区网络; 敏感特征; 谱密度; 融合聚类

Clustering algorithm of sensitive information in community network based on block detection

WANG Xuejun

(Guangzhou Huali College, Guangzhou 511325, China)

[Abstract] In order to improve the clustering ability of sensitive feature information of wireless virtual community network, it is necessary to carry out data optimization clustering processing. This paper proposes a grid strengthening clustering algorithm for sensitive feature information of wireless virtual community network based on block text similarity detection. The heterogeneous directed graph analysis method is used to design the storage structure of sensitive feature information of wireless virtual community network, after that the feature space reorganization technology is used to reorganize the sensitive feature information structure of wireless virtual community network, and the related information feature quantity of sensitive feature information of wireless virtual community network is extracted. The block text similarity detection method is used to realize the spectral density feature extraction and fusion clustering processing of sensitive feature information of community network. The simulation results show that the clustering of sensitive feature information spectral density fusion of community network by this method is good, and the ability of block detection of sensitive information of community network is strong.

[Key words] community network; sensitive features; spectral density; fusion clustering

0 引言

随着虚拟网络和社区网络信息传播数量的不断增强,社区网络的多维度联动、多主体参与性特征越来越明显,结合运行机制及危机治理模式化运行,需要建立社区网络敏感特征信息检测模型,通过探究影响网络舆情热度的主要因素,采用主题词和敏感词特征分析方法,实现对无线虚拟社区网络敏感性的影响因素和情感分析,提高无线虚拟社区网络敏感信息检测能力^[1]。

当前,对无线虚拟社区网络敏感特征信息聚类方法主要有 K-Means 聚类法、社区网络敏感区域模糊聚类方法、智能仿生群聚类方法等,建立无线虚拟社区网络敏感特征信息的特征提取和大数据融合模型^[2],通过相应的融合聚类算法,实现对无线虚拟

社区网络敏感特征检测。文献[3]中提出基于权重差异度的动态模糊聚类算法,采用加权自适应学习实现大数据聚类,但该方法的计算开销较大,实时性不好。文献[4]中提出优化子空间的高维聚类算法,通过社区网络敏感特征信息分块层次结构设计,提高数据聚类的精度,但该方法的模糊度较大,特征聚类性不好。针对上述问题,本文提出基于分块文本相似度检测的无线虚拟社区网络敏感特征信息网格强化聚类算法。首先采用异构有向图分析方法进行无线虚拟社区网络敏感特征信息存储结构设计,然后采用分块文本相似度检测的方法实现对社区网络敏感特征信息谱密度特征提取和融合聚类处理。最后进行仿真测试分析,展示了本文方法在提高无线虚拟社区网络敏感特征信息网格强化聚类能力方面的优越性能。

作者简介: 王学军(1973-),男,硕士,讲师,主要研究方向:社区发现、聚类算法研究。

通讯作者: 王学军 Email: wxjgdut@163.com

收稿日期: 2022-05-18

1 敏感特征信息结构分析和特征提取

1.1 无线虚拟社区网络敏感特征信息结构分析

采用有向图模型构建无线虚拟社区网络敏感特征信息的聚类节点分布结构模型,引入知识图谱提升推荐方法,得到无线虚拟社区网络敏感特征信息的异构存储结构模型^[5],计算无线虚拟社区网络敏感特征信息的模糊聚类特征分布集,得到融合度函数定义为:

$$d_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (1)$$

其中, x_i, y_i, z_i 分别表示敏感特征参数分布集,采用语义本体模型构造的方法,进行无线虚拟社区网络敏感特征信息表达,分析无线虚拟社区网络敏感特征信息集统计差异分布特性,采用分段线性拟合方法进行无线虚拟社区网络敏感特征信息谱聚类,得到每个社区网络敏感特征插值点的实际值为:

$$d_{m+1}(m) = d_{k+1}(m) \pm \sqrt{(d_m(0)e^{\lambda_1} + 1)^2 - \sum_{i=1}^{m-1} [d_{m+1}(i) - d_{k+1}(i)]^2} \quad (2)$$

其中: $d_{m+1}(m)$ 为无线虚拟社区网络敏感特征信息集在第 m 点的预测值, $d_{k+1}(m)$ 为采用第 m 点处采集的无线虚拟社区网络敏感特征信息的模糊性特征量。根据无线虚拟社区网络敏感特征信息的特征提取结果^[6],采用回归分析方法建立无线虚拟社区网络敏感特征信息聚类模型,得到模糊回归分布系数为:

$$Mi = L_m + \frac{N \times 0.5 - \sum f_{less}}{f_m} \times width \quad (3)$$

其中, Mi 表示无线虚拟社区网络敏感特征信息聚类的中位数; L_m 为无线虚拟社区网络敏感特征信息聚类的下界; f_m 为无线虚拟社区网络敏感特征信息的中位数; f_{less} 表示各维度下无线虚拟社区网络敏感特征信息的最小统计特征量。

1.2 网络敏感特征信息特征提取

分析无线虚拟社区网络敏感特征信息的自相关特征量,通过对数据的模糊聚类和线性规划设计,得到无线虚拟社区网络敏感特征信息的状态特征分布集为 $p(x_0)$,构建资源聚类的节点分布模型,在大数据背景下进行无线虚拟社区网络敏感特征信息网格强化聚类优化设计^[7],假设无线虚拟社区网络敏感特征信息聚类节点图模型属性集为 $X = \{x_1, x_2, \dots,$

$x_n\}$,采用语义本体模型构造的方法,进行无线虚拟社区网络敏感特征信息分析,通过自相关特征匹配方法,求得无线虚拟社区网络敏感特征信息的关联规则概念集,对无线虚拟社区网络敏感特征信息进行粗糙集调度,得到对贴近度 φ_1 定义为:

$$(\bar{s}, \bar{a}) = \varphi_1((s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)) = \Delta(\sum_{j=1}^n \omega_j \Delta^{-1}(s_j, a_j)) \quad (4)$$

其中, $\sum_{j=1}^n \omega_j = 1, \bar{s} \in S, \bar{a} \in [-0.5, 0.5]$ 。

设 $\{(s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)\}$ 是一组描述无线虚拟社区网络敏感特征信息聚类的语义特征分布集,建立无线虚拟社区网络敏感特征信息聚类的统计分析模型,在模糊网格区域,进行无线虚拟社区网络敏感特征信息聚类的统计分析和自适应调度,采用多尺度一维卷积小波分析方法进行无线虚拟社区网络敏感特征信息采样和融合处理^[8]。

2 信息融合聚类

2.1 敏感特征信息结构重组

采用网格化的数据采样,得到无线虚拟社区网络敏感特征信息标量时间序列为 $x(t)$,并且 $t = 0, 1, \dots, n - 1$,给定无线虚拟社区网络敏感特征信息信息流的一向量组 $x_1, x_2, \dots, x_n \in C^m$ (m 维复数空间),将无线虚拟社区网络敏感特征信息聚类的特征提取问题转化为一个二元语义决策问题,无线虚拟社区网络敏感特征信息聚类的模糊特征匹配评价指标集为 $E_k \in E(k = 1, 2, \dots, t)$,采用主题词匹配的方法,分析 X 的相似度函数,得到无线虚拟社区网络敏感特征信息的模糊隶属度函数为:

$$\tilde{w}_k = \tilde{w}_{k-1} \frac{p(z_k / \tilde{x}_k^i) p(\tilde{x}_k^i / x_{k-1}^i)}{q(\tilde{x}_k^i / x_{k-1}^i)} \quad (5)$$

其中, \tilde{x}_k^i 为无线虚拟社区网络敏感特征信息的语义特征量, \tilde{w}_{k-1}^i 为惯性权重。利用概念、实例和属性等实体集,得到无线虚拟社区网络敏感特征信息聚类中心节点在 $k + 1$ 时刻的迭代函数为:

$$x_i(k+1) = x_i(k) + s \left(\frac{x_j(k) - x_i(k)}{\|x_j(k) - x_i(k)\|} \right) \quad (6)$$

其中, $x_j(k), x_i(k)$ 分别表示社区网络敏感特征信息聚类参数,采用决策树模型进行无线虚拟社区网络敏感特征信息的特征重构,得到无线虚拟社区网络敏感特征信息的模糊信息加权权重向量 v_i ,无线虚拟社区网络敏感特征信息的相关特征分布矩阵表示为:

$$U = \{\mu_{ik} \mid i = 1, 2, \dots, c, k = 1, 2, \dots, n\} \quad (7)$$

其中, c 为无线虚拟社区网络敏感特征信息聚类的搜索步数, μ_{ik} 为无线虚拟社区网络敏感特征信息的语义关联度决策系数。根据上述分析, 进行无线虚拟社区网络敏感特征信息的特征分析和优化调度, 得到无线虚拟社区网络敏感特征信息的模糊聚类中心为 d_i , 在融合聚类中心满足 $m \rightarrow 1$ 时, $sn\xi \rightarrow \tanh\xi$, 采用分块特征演化方法, 进行无线虚拟社区网络敏感特征信息的关联特征检测, 结合模糊相关性融合的方法, 进行无线虚拟社区网络敏感特征信息的网格强化聚类分析^[9-11]。

2.2 谱密度聚类分析

采用模糊 C 均值聚类方法进行无线虚拟社区网络敏感特征信息的网格强化聚类和属性分类识别, 结合特征空间重组技术进行无线虚拟社区网络敏感特征信息结构重组, 提取无线虚拟社区网络敏感特征信息的关联信息特征量, 得到无线虚拟社区网络敏感特征信息的在线聚类准则为:

$$S(i, j) = \frac{\sum_{u \in U_{ij}} (V_{u,i} - 3)(V_{u,j} - 3)}{\sqrt{\sum_{u \in U_{ij}} (V_{u,i} - \bar{V}_i)^2} \sqrt{\sum_{u \in U_{ij}} (V_{u,j} - \bar{V}_j)^2}} \quad (8)$$

无线虚拟社区网络敏感特征信息聚类节点的深度学习加权系数为 $W_e = (\omega_j^{(e)}, 0)$ 。修正每个无线虚拟社区网络敏感特征信息聚类自适应加权学习系数 v_i , 得到无线虚拟社区网络敏感特征信息聚类有效性评价矩阵 $R = (r_{ij}, a_{ij})_{m \times n}$ 和指标权重 $W = ((\omega_1, \beta_1), (\omega_2, \beta_2), \dots, (\omega_n, \beta_n))$, 通过谱密度聚类, 得到无线虚拟社区网络敏感特征信息的约束规划模型为:

$$\min(f) = \sum_{i=1}^m \sum_{j=1}^n C_{ij} X_{ij} \quad (9)$$

$$\begin{cases} \sum_{j=1}^n X_{ij} = a_i & i = 1, 2, \dots, m \\ \sum_{i=1}^m X_{ij} = b_j & j = 1, 2, \dots, n \\ X_{ij} \geq 0 & i = 1, 2, \dots, m, j = 1, 2, \dots, n \end{cases} \quad (10)$$

此时, 无线虚拟社区网络敏感特征信息聚类的最优评价集记为 L_1, \dots, L_n 和 $P_1^{\min}, \dots, P_n^{\min}$, 根据稀疏化特征值分析无线虚拟社区网络敏感特征信息的传输码元, 当码元个数为 N_f 个, 无线虚拟社区网络敏感特征信息谱采样时间间隔为 T_f , 无线虚拟社区

网络敏感特征信息的优化聚类模型为:

$$l_{d_{ij} \rightarrow c_x} = \frac{\sum_{v=0}^{|c_x|} \cosin_{ij \rightarrow x}(d_{ij}, d_{xv}) \ddot{\circ}^{-1}}{|c_x| \ddot{\circ}} \quad (11)$$

其中, $\cosin_{ij \rightarrow x}(d_{ij}, d_{xv})$ 为无线虚拟社区网络敏感特征信息的融合聚类特征集, 综上所述, 实现对社区网络敏感特征信息谱密度融合聚类处理^[12]。

3 仿真测试

采用 C++ 和 Matlab 7 混合编程进行无线虚拟社区网络敏感特征信息聚类的算法处理, 在 Hadoop 云平台中构建无线虚拟社区网络敏感特征信息数据库结构模型, 无线虚拟社区网络敏感特征信息大数据分布的初始样本规模为 1 200, 模糊系数 m 设为 2, 社区网络敏感特征采样集的相似度为 0.38, 根据上述仿真参数设定, 进行无线虚拟社区网络敏感特征信息采样, 无线虚拟社区网络敏感特征信息的时域分布如图 1 所示。

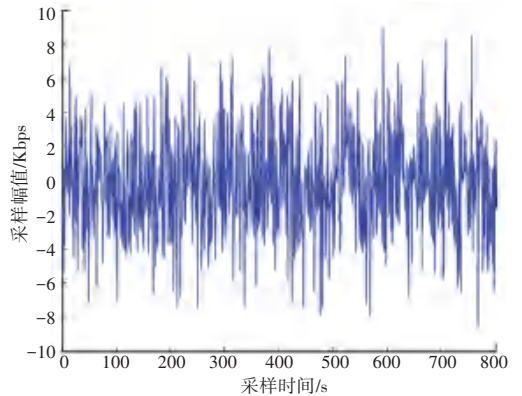
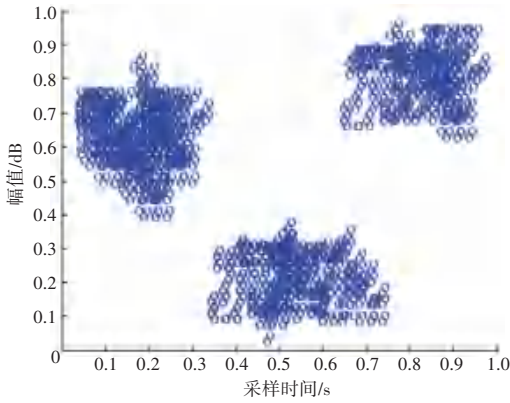


图 1 无社区网络敏感特征信息数据的时域波形

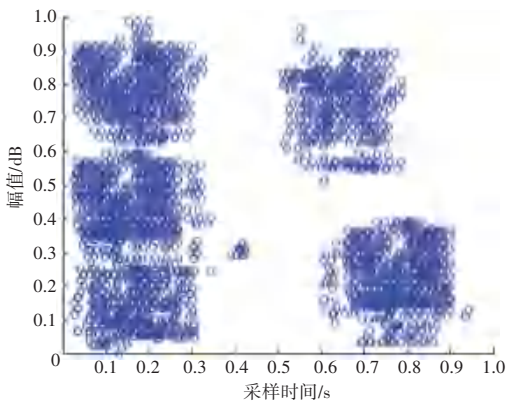
Fig. 1 Time domain waveform of sensitive characteristic information data of community-free network

这里, 以图 1 中的数据为研究对象, 对提取的无线虚拟社区网络敏感特征信息特征量进行模糊聚类处理, 提取无线虚拟社区网络敏感特征信息的谱特征量, 实现网格强化聚类, 得到数据聚类输出如图 2 所示。

分析图 2 仿真结果得知, 本文方法能有效实现对无线虚拟社区网络敏感特征信息融合聚类处理, 测试对社区网络敏感信息的分块检测准确性, 得到误分率对比结果如图 3 所示。分析图 3 得知, 采用该方法进行社区网络敏感特征信息谱密度融合的聚类性较好, 对社区网络敏感信息的分块检测能力较强。



(a) 测试样本



(b) 训练样本

图 2 数据聚类输出

Fig. 2 Data clustering output

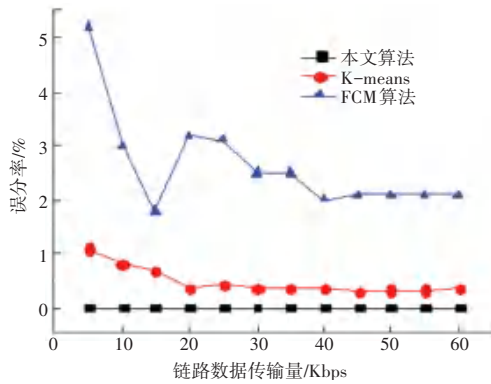


图 3 误分率对比

Fig. 3 Comparison of error rate

4 结束语

通过对虚拟社区网络敏感特征信息融合聚类处理,提高对虚拟社区网络敏感特征信息检测能力,建立社区网络敏感特征信息检测模型,本文提出基于分块文本相似度检测的无线虚拟社区网络敏感特征信息网格强化聚类算法,分析无线虚拟社区网络敏感特征信息集统计差异分布特性,采用分段线性拟合方法进行无线虚拟社区网络敏感特征信息谱聚类,采用分块特征演化方法,进行无线虚拟社区网络敏感特征信息的关联特征检测,结合模糊相关性融合的方法,实现对信息的谱密度融合聚类。分析得知,本文方法对社区网络敏感特征信息谱密度融合的聚类性较好,提高了社区网络敏感信息的分块检测能力。

参考文献

- [1] 魏新艳,张琳. 物联网中基于信任的频谱资源分配机制[J]. 计算机工程, 2020, 46(04): 26-32, 39.
- [2] 王俊红, 闫家荣. 基于欠采样和代价敏感的不平衡数据分类算法[J]. 计算机应用, 2021, 41(01): 48-52.
- [3] 刘良凤, 刘三阳. 基于权重差异度的动态模糊聚类算法[J]. 吉林大学学报(理学版), 2019, 57(03): 574-582.
- [4] 吴涛, 陈黎飞, 郭躬德. 优化子空间的高维聚类算法[J]. 计算机应用, 2014, 34(08): 2279-2284.
- [5] 刘学文, 王继奎, 杨正国, 等. 密度峰值优化的球簇划分欠采样不平衡数据分类算法[J]. 计算机应用, 2022, 42(05): 1455-1463.
- [6] 李艳霞, 柴毅, 胡友强, 等. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(04): 673-688.
- [7] 刘治纲, 肖庆汇, 丁雪非, 等. 软件定义网络用户动态访问控制模型仿真[J]. 计算机仿真, 2019, 36(01): 308-311, 396.
- [8] 聂茜婵, 张阳, 余敦辉, 等. 面向全局优化的时空众包任务分配算法[J]. 计算机应用, 2020, 40(07): 1950-1958.
- [9] 宋天舒, 童咏昕, 王立斌, 等. 空间众包环境下的3类对象在线任务分配[J]. 软件学报, 2017, 28(03): 611-630.
- [10] 胡甜甜, 但雅波, 胡杰, 等. 基于注意力机制的Bi-LSTM结合CRF的新闻命名实体识别及其情感分类[J]. 计算机应用, 2020, 40(07): 1879-1883.
- [11] 刘璟. 中文命名实体识别方法研究[J]. 电脑知识与技术, 2019, 15(09): 179-180.
- [12] 方军, 张璋, 张雪峰, 等. 基于均衡适应度的云 workflow 调度算法[J]. 计算机应用与软件, 2019, 36(05): 255-261.