

文章编号: 2095-2163(2019)03-0176-04

中图分类号: TP393

文献标志码: A

# 基于模糊分区聚类的社交网络用户情景模式预测

张创基

(广州华立科技职业学院, 广州 511325)

**摘要:** 为了提高社交网络用户行为分析和情景模式预测能力,优化社交网络建设,结合数据挖掘和行为分析方法进行社交网络的用户特征分析和用户情景模式的优化挖掘,发现社交网络用户行为特征。提出一种基于模糊分区聚类的社交网络用户情景模式预测方法,构建社交网络用户情景模式分布的关联拓扑结构模型,采用 Parallel Sets 变元轴排序方法进行社交网络用户情景模式存储结构分区调度,结合分段特征提取方法进行社交网络用户情景模式的关联数据挖掘,采用自适应寻优方法求取社交网络用户的情景模式的分布信息,利用模糊分区聚类方法发现用户情景模式数据集中的隐含模式,根据数据模糊分区聚类 and 挖掘结果,实现社交网络用户情景模式的自适应预测。仿真结果表明,采用该方法进行社交网络用户情景模式预测的准确性较高,提高了对社交网络用户情景模式特征配准的精度,算法处理的实时性较好。

**关键词:** 数据聚类; 社交网络; 用户情景模式; 特征提取

## Prediction of user scenario patterns in social networks based on fuzzy partition clustering

ZHANG Chuangji

(Guangzhou Huali Science and Technology Vocational College, Guangzhou 511325, China)

**[Abstract]** In order to improve the ability of social network user behavior analysis and scenario pattern prediction, and optimize the construction of social network, combining the methods of data mining and behavior analysis, the user characteristics analysis and user scenario pattern optimization mining of social network are carried out. The behavior characteristics of social network users could be discovered. In this paper, a prediction method of social network user scenario pattern based on fuzzy partition clustering is proposed, and the associated topology model of social network user scenario pattern distribution is constructed. The Parallel Sets argument axis sorting method is used to implement the partition scheduling of the storage structure of the social network user scenario pattern, and the segmented feature extraction method is used to mine the associated data of the social network user scenario pattern. The adaptive optimization method is used to obtain the distribution information of the social network user scenario pattern, the fuzzy partition clustering method is used to find the hidden patterns in the user scenario pattern data set, and according to the data fuzzy partition clustering and mining results, the adaptive prediction of social network user scenarios is realized. The simulation results show that the proposed method has a high accuracy in predicting social network user scenarios, improves the accuracy of feature registration of social network user scenarios, and the algorithm can deal with the real-time performance better.

**[Key words]** data clustering; social networks; user scenarios; feature extraction

## 0 引言

随着 Web2.0 快速的发展,社交网络得到快速发展,社交网络逐渐变成人们情感交流、关系维护和信息沟通的主要平台和方式。社交网络为用户建立在线社交关系,通过互动交流实现对具有共同特征爱好的用户聚类处理,在不同的情景模式下构建社交网络模块,提高信息资源的自适应分配和调度能力。在社交网络平台上,需要对用户的情景模式进行自动预测,结合大数据分析,进行用户的行为特征分析和用户的关系分析,根据用户的社交搜索和偏好,实现情景模式预测和调度,提高社交网络的服务功能,结合数据挖掘和行为分析方法进行社交

网络的用户特征分析和情景模式预测,发现社交网络用户的规律性<sup>[1]</sup>。研究社交网络用户情景模式预测方法在社交网络的构建和行为特征分析中具有重要意义。

当前,常见的聚类方法有 K-means 算法、模糊 C 均值算法和层次分割聚类挖掘方法<sup>[2-3]</sup>。模糊 C 均值聚类方法在进行社交网络用户情景模式数据挖掘的聚类处理中容易陷入局部最优解,而层次分割聚类方法进行大数据关联挖掘受到分割阈值的影响<sup>[4]</sup>,导致对初始的聚类中心具有较大的敏感性,针对上述问题,本文提出一种基于模糊分区聚类的社交网络用户情景模式预测方法,构建社交网络用户情景模式分布的关联拓扑结构模型,采用分段特

**作者简介:** 张创基(1983-),男,硕士,讲师,网络工程师,主要研究方向:信息与网络安全、计算机控制、数据挖掘。

**收稿日期:** 2018-12-20

征提取方法进行社交网络用户情景模式的关联特征提取,采用模糊 C 均值聚类方法进行数据挖掘,实现社交网络用户情景模式预测。最后通过仿真实验进行性能测试,展示本文方法在提高用户情景模式预测准确性方面的优越性能。

## 1 社交网络用户情景模式分布的关联拓扑结构模型与特征提取预处理

### 1.1 社交网络用户情景模式拓扑结构模型

社交网络信息资源传输系统设计中,用矩阵  $C$  表示社交网络的用户拓扑结构,其中,  $C$  是  $N \times N$  的二维矩阵。 $N$  为在网络中的节点个数。对于用户节点  $i$ , 假设  $N_i^1$  为第一个相邻节点, 得到社交网络用户情景模式分布的多元邻节点为:

$$N_i^2 = N_i^1 \cup \left( \bigcup_{j \in N_i^1} N_j^1 \right), \quad (1)$$

假设由  $M$  个时隙组成了一个帧,用矩阵  $X$  表达时隙的分配,而  $X = x_{mi}$  是  $M \times N$  维数的矩阵。其中,当社交网络用户情景模式分布节点  $i$  在第  $m$  个时隙发送数据时,  $x_{mi}$  的取值为 1, 否则  $x_{mi}$  的取值为 0, 在不同的初始聚类中心<sup>[5]</sup>, 社交网络用户情景模式存储的属性分布用子图  $G_1 = (M^{\alpha_1}, M^{\beta_1}, Y_1), G_2 = (M^{\alpha_2}, M^{\beta_2}, Y_2)$  表示, 则满足  $G_1 \subseteq G_2 \Leftrightarrow Y_1 \subseteq Y_2$ , 令  $A = \{a_1, a_2, \dots, a_n\}$  为用户轨迹分布的幅值, 社交网络用户情景模式存储节点的状态特征用  $P(i, j)$  表示,  $p(i, j) = \lim_{t \rightarrow \infty} p\{a_t = i, b_t = j\}$ , 那么社交网络用户情景模式分布的拓扑树的总平均子节点数为  $M$  层子节点数总和与层次总和的比值, 即:

$$\bar{D} = \sum_{i=1}^{M-1} |D_{i_i}| / \sum_{j=1}^{M-1} |L_j|. \quad (2)$$

将数据用户节点  $\alpha(i, j)$  输入数据到链路层中, 得到关键路径节点采集的社交网络用户情景模式时间序列为  $x_n$ , 期望响应为  $d_n$ , 由此构建社交网络的用户情景模式分布轨迹拓扑结构图如图 1 所示。

### 1.2 用户情景模式跟踪的特征提取

在构建社交网络用户情景模式分布的关联拓扑结构模型的基础上, 进行用户情景模式的特征提取, 采用分段特征提取方法进行社交网络用户情景模式的关联特征提取<sup>[6]</sup>, 用户情景模式传输树中任意一个非根节点的情景模式社交网络用户情景模式预测的步进值为:

$$CW_{\min}^{l+1} = \begin{cases} CW_{\min}^0 \times (1 + \bar{D}_0) \chi, & i = 0; \\ CW_{\min}^l \times (1 + \bar{D}_i) \chi, & 0 < i < M. \end{cases} \quad (3)$$

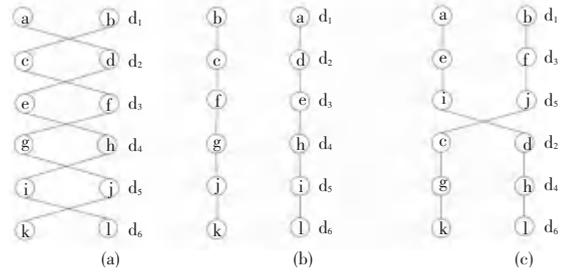


图 1 社交网络用户情景模式拓扑结构

Fig. 1 Topology of social network user scenario pattern

社交网络用户情景模式预测的 QoS 控制加权自适应学习系数表示为:

$$\chi = \log_{(1+\bar{D})} M A / CW_{\min}^0, \quad (4)$$

其中,  $CW_{\min}^0$  为社交网络用户情景模式拓扑结构传输树中最底层根节点的预测初始化值; 而  $CW_{\min}^l$  为第  $l_i$  步的社交网络用户情景模式拓扑控制值, 在情景模式预测过程中, 每一步之间加权控制构成是相同的, 假设混合云环境下社交网络用户情景模式簇头节点上的信息码表达为:

$$u_{sw} = -K \text{sat}(S(t)/\mu) / (\lambda g_x + g_\theta). \quad (5)$$

设  $U$  是一个用精确数值表示的定量论域,  $C$  是  $U$  上的定性概念, 采用无向图模型结构表示原始感知的社交网络用户情景模式, 对社交网络用户情景模式进行自适应调度中, 融合模型由一个参数未知多重假设检验问题构成<sup>[7]</sup>, 采用 Parallel Sets 变元轴排序方法进行社交网络用户情景模式存储结构分区调度, 调度模型为:

$$\begin{cases} \max U = u_1 + u_2 + \dots + u_n; \\ u_i = p_i; \\ \sum_{i=1}^n p_i = 1, 0 < p_i < 1; \\ \frac{p_1 / (1 - p_1)}{w_1} = \frac{p_i / (1 - p_i)}{w_i} = \dots = \\ \frac{p_n / (1 - p_n)}{w_n} = \frac{1}{K}. \end{cases} \quad (6)$$

其中,  $K$  表示加权调节系数, 根据社交网络中用户层和层之间的平均子节点数目, 进行特征提取和分布式的网络组网设计<sup>[8]</sup>, 此时社交网络用户情景模式多源节点中形成新的映射, 可得:

$$x_n = [x(0), x(1), \dots, x(N-1)]^T. \quad (7)$$

根据社交网络用户情景模式的分区信息熵的信息流特征, 进行特征分解, 定义特征值的受限条件关联规则挖掘量为:

$$\text{cum}(\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_k x_k) = \left( \prod_{i=1}^k \lambda_i \right) \text{cum}(x_1, x_2, \dots, x_k). \quad (8)$$

其中,该过程的 $k$ 阶累积量 $c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1})$ ,提取任务信息流的特征尺度,在社交网络用户情景模式分区自适应调度的受限条件下,提取关联规则特征量,为社交网络的用户情景模式跟踪提供数据输入基础。

## 2 预测算法优化

### 2.1 社交网络用户情景模式的关联数据挖掘

在构建社交网络用户情景模式分布的关联拓扑结构模型和特征提取的基础上,进行用户情景模式预测算法的改进设计,本文提出一种基于模糊分区聚类的社交网络用户情景模式预测方法,采用分段特征提取方法进行社交网络用户情景模式的关联数据挖掘,根据用户情景模式特征取小取大后的比值描述社交网络用户情景模式的相似程度,得到社交网络用户情景模式的互信息特征的熵函数可以表示为:

$$p_i = f(d_i, d'_i) / N = T_i, O_{j_{f_1}}, \dots, O_{j_{f_n}}, T_i, \dots, T_i, O_{n_{f_1}}, \dots, O_{n_{f_n}}, \quad (9)$$

社交网络用户情景模式集合基数满足 $p_0 = \frac{1}{m}$ ,  
 $m = r_1 + r_2 + \dots + r_i + \dots + r_n$ ,优化的社交网络用户情景模式关联规则特征提取为:

$$p_i = \frac{r_i}{m} = \frac{r_i w_0}{m w_i} = \frac{w_i}{w_1 + \dots + w_i + \dots + w_n}, \quad (10)$$

当簇内相邻节点身份的特征值满足: $A_1 \cup A_2 \cup \dots \cup A_k = A$ ,且 $A_i \cap A_j = \Omega$ ,其中 $i, j = 1, \dots, m$ 且 $i \neq j$ ,得到单个社交网络用户节点 $i$ 的信息熵:

$$H_i(x) = \sum_{k=1}^K p_k \ln \frac{1}{p_k} = - \sum_{k=1}^K p_k \ln p_k, \quad (11)$$

采用自适应寻优对 $H_i(x)$ 求取社交网络用户的情景模式的分布信息最大值 $\max(H_i(x))$ ,由此得到社交网络用户情景模式的熵融合特征提取最优化的约束条件为:

$$\begin{aligned} & \uparrow f_1 < f_2, \text{ if } \mu_1, \mu_2 \geq \alpha; \\ (f_1, \mu_1) <_{\alpha} (f_2, \mu_2) & \Leftrightarrow \uparrow f_1 < f_2, \text{ if } \mu_1 = \mu_2; \\ & \uparrow \mu_1 > \mu_2, \text{ otherwise.} \end{aligned} \quad (12)$$

其中, $f_2, \mu_1, \mu_2$ 分别表示社交网络用户情景模式相似度特征 $\vec{x}_1$ 和 $\vec{x}_2$ 的目标函数值。采用 Parallel Sets 变元轴排序方法进行社交网络用户情景模式排

序,根据数据挖掘结果进行轨迹跟踪聚类分析。

### 2.2 模糊聚类及用户情景模式预测实现

结合模糊分区聚类方法发现数据集中的隐含模式,得到用户情景模式挖掘的异步递进加权系数为:

$$CW_{\min}^n = \begin{cases} ((1 - B_{l(n)})e^{1-\alpha_n} + B_{l(n)}) \times CW_{\min}^{l(n)}, & \alpha_n > 1; \\ CW_{\min}^{l(n)}, & 0 \leq \alpha_n \leq 1. \end{cases} \quad (13)$$

$$B_{l(n)} = (1 / (1 + \bar{D}_{l(n)-1}))^{\chi}, \quad (14)$$

$$\alpha_n = \begin{cases} |D_n| / (\bar{D}_{l(n)}), & \bar{D}_{l(n)} \neq 0; \\ 0, & \bar{D}_{l(n)} = 0. \end{cases} \quad (15)$$

其中, $\alpha_n$ 为子节点异步迭代系数, $B_{l(n)}$ 为网络用户轨迹跟踪的调整系数。同时,对于任意网络用户节点 $n$ ,由于:

$$(1 + \bar{D})^M > 1, \quad (16)$$

$$A / CW_{\min}^0 > 1, \quad (17)$$

因此,得社交用户情景模式预测的模糊分区聚类中心调整系数为:

$$\chi > 0, \quad (18)$$

由 $(1 + \bar{D}_i)^{\chi} > 1$ ,用户情景模式数据挖掘的输出特征值满足:

$$CW_{\min}^{l_{i+1}} > CW_{\min}^{l_i} \quad (19)$$

综上所述,利用模糊分区聚类方法发现用户情景模式数据集中的隐含模式,根据数据模糊分区聚类和挖掘结果,实现社交网络用户情景模式的自适应预测。

## 3 仿真实验与结果分析

为了测试本文方法在实现社交网络用户情景模式预测中的性能,进行仿真实验,实验的软件环境为 Matlab 7,社交网络的用户节点数为 100,分区的大小为 1 000M,用户的分布层数为 12,嵌入维数为 3,关联系数 0.23,对用户轨迹位置采样的离散采样率为 $f_s = 10 * f_0$  Hz = 10 KHz,训练样本规模为 12 000,根据上述仿真环境和参数设定,进行社交网络的用户情景模式分布的关联数据挖掘,得到挖掘结果如图 2 所示。

根据图 2 的数据挖掘结果进行用户情景模式预测,得到预测输出如图 3 所示。

分析图 3 得知,采用本文方法进行社交网络的用户情景模式预测具有很好的抗干扰能力,特征聚类性较好。采用不同方法进行预测精度对比,得到

对比结果如图4所示,分析图4得知,采用本文方法进行社交网络的用户情景模式预测的精度较高。

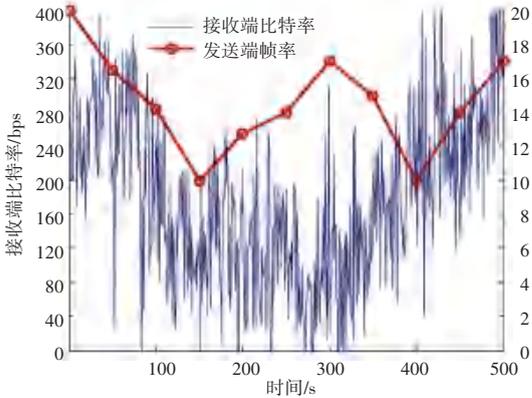


图2 社交网络的用户情景模式分布的关联数据挖掘

Fig. 2 Associated data mining for user scenario pattern distribution in social networks

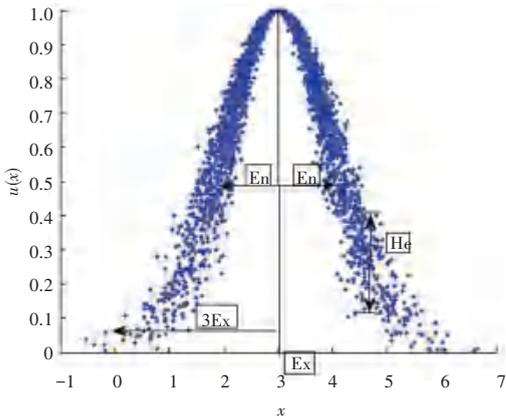


图3 情景模式的大数据特征预测结果输出

Fig. 3 Output of big data feature prediction result of scenario model

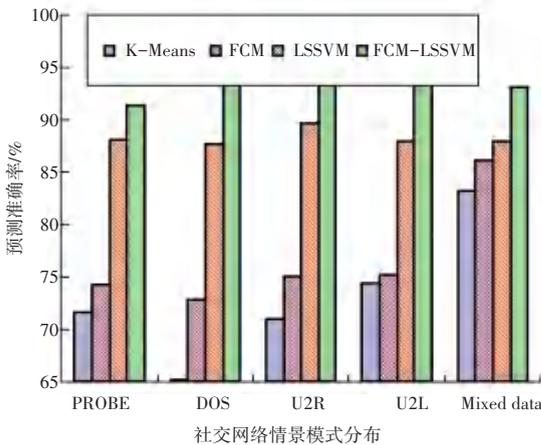


图4 预测准确性对比

Fig. 4 Prediction accuracy comparison

### 4 结束语

对用户的情景模式进行自动预测,结合大数据分析,进行用户的行为特征分析和用户的关系分析,提高社交网络的服务功能,本文提出一种基于模糊分区聚类的社交网络用户情景模式预测方法,构建社交网络用户情景模式分布的关联拓扑结构模型,采用分段特征提取方法进行社交网络用户情景模式的关联特征提取,采用 Parallel Sets 变元轴排序方法进行社交网络用户情景模式存储结构分区调度,结合模糊分区聚类方法发现数据集中的隐含模式,采用模糊 C 均值聚类方法进行数据挖掘,实现社交网络用户情景模式预测。研究得知,采用本文方法进行社交网络用户情景模式预测的准确性较高,数据挖掘精度较高,在社交网络用户情景模式挖掘和行为分析中具有很好的实践价值。

### 参考文献

- [1] 曹玖新,董丹,徐顺,等. 一种基于 k-核的社会网络影响最大化算法[J]. 计算机学报,2015,38(2):238-248.
- [2] 曹玖新,陈高君,吴江林,等. 基于多维特征分析的社交网络意见领袖挖掘[J]. 电子学报,2016,44(4):898-905.
- [3] 刘亚州,王静,潘晓中,等. 社交网络中考虑节点度的演化博弈[J]. 计算机应用,2018,38(4):1029-1035.
- [4] 朱冠桦,蒋国平,夏玲玲. 社交网络上从众现象对谣言传播影响的研究[J]. 计算机科学,2016,43(2):135-139,143.
- [5] 张付志,刘赛,李忠华,等. 融合用户评论和环境信息的协同过滤推荐算法[J]. 小型微型计算机系统,2014,35(2):228-232.
- [6] 郭磊,马军,陈竹敏,等. 一种结合推荐对象间关联关系的社会化推荐算法[J]. 计算机学报,2014,37(1):219-228.
- [7] 孟祥武,刘树栋,张玉洁,等. 社会化推荐系统研究[J]. 软件学报,2015,26(6):1356-1372.
- [8] 吴铮,于洪涛,刘树新,等. 基于信息熵的跨社交网络用户身份识别方法[J]. 计算机应用,2017,37(8):2374-2380.