

doi:10.16055/j.issn.1672-058X.2021.0005.019

# 基于逻辑回归的企业二次创业金融数据 分类方法研究\*

赖红清

(佛山职业技术学院 工商管理学院商贸系,广东 佛山 528200)

**摘要:**企业二次创业金融数据的优化分类能提高数据的统计分析能力,提出基于逻辑回归的企业二次创业金融数据分类方法,采用自适应无监督学习的方法进行数据统计的融合处理,构建数据分布的不规则空间聚类模型,采用相空间结构重组方法进行数据的模糊特征重构,提取企业二次创业金融数据的关联规则特征量,采用逻辑回归分析方法进行数据分类的融合聚类处理,结合模糊C均值聚类方法,实现对数据分类的自适应寻优和收敛性控制,实现数据分类优化。仿真结果表明:采用该方法进行企业二次创业金融数据分类的准确性较高,收敛性较好,特征聚类能力较强。

**关键词:**逻辑回归;企业二次创业;金融数据;分类

**中图分类号:**F832

**文献标志码:**A

**文章编号:**1672-058X(2021)05-0114-06

## 0 引言

随着经济的发展,需要对企业二次创业金融数据进行优化聚类分析,通过提取企业二次创业金融数据的统计特征量,结合大数据信息处理和数据融合方法,将其进行优化聚类分析,以提高企业二次创业金融数据的模糊聚类能力<sup>[1]</sup>。对企业二次创业金融数据的分类是实现企业金融管理优化的关键,相关的数据分类研究受到人们的极大关注<sup>[2]</sup>。对企业二次创业金融数据的聚类分析主要采用模糊聚类方法,结合统计分析,进行数据分类处理,但传统方法模糊度较大,分类的准确性不好。针对上

述问题,提出基于逻辑回归的企业二次创业金融数据分类方法,采用逻辑回归分析方法将数据进行融合聚类处理,结合模糊C均值聚类方法,实现数据分类的自适应寻优和收敛性控制。最后进行仿真实验分析,展示了该方法在提高企业二次创业金融数据分类能力方面的优越性能。

## 1 企业二次创业金融数据分布模型及特征分析

### 1.1 分布模型

为了实现企业二次创业金融数据分类,采用自

收稿日期:2019-10-09;修回日期:2020-06-15.

\* 基金项目:2018年佛山市哲社规划课题(2018-GJ071).

作者简介:赖红清(1973—),女,广东省茂名市人,硕士研究生,副教授,从事企业管理、产业经济和技术创新研究.

适应无监督学习方法进行统计融合处理。首先构建企业二次创业金融数据模糊决策树模型,采用相空间重构方法进行模糊特征重构<sup>[3]</sup>,采用一个四元组  $(E_i, E_j, d, t)$  来表示特征分布权系数,其中  $E_i, E_j$  是企业二次创业金融数据特征权重的实体集(即节点  $i$  和  $j$ ),  $d$  为特征权重的交互性统计数据,  $t$  为企业二次创业过程中的业绩关联数据特征集,采用企业内部代理方法进行统计特征分析,得到企业二次创业金融数据的量化集为

$$\mathbf{x}(t) = (x_0(t), x_1(t), \dots, x_{k-1}(t))^T \quad (1)$$

采用一个  $1 \times N$  维的分布阵列进行企业二次创业金融数据特征重构,确定模糊时间窗口值  $N$ ,构建多维信息熵分布矩阵。采用激励机制建立企业二次创业金融数据的特征权重分析模型,建立窄时域窗  $TL_x, TL_y$ ,得到企业二次创业金融数据特征权重的模糊特征提取模型为

$$TL(x, y) = \begin{cases} Text, & \text{if } (D_x(x, y) \leq N) \\ NonText, & \text{Otherwise} \end{cases} \quad (2)$$

式(2)中,  $D_x(x, y)$  表示特征提取最大范围值,当提范围值小于等于窗口值  $N$  时,进行特征权重的模糊特征提取;当超出窗口值  $N$  时,不进行提取。

设企业二次创业金融数据特征权重的模板特征分布为  $m$ ,构建 Probit 多元回归分析模型<sup>[4]</sup>,得到数据的有效性控制指数为  $N_{j*}$ ,在进行企业二次创业金融数据特征权重回归分析的基础上,得到企业二次创业金融数据的统计回归分布为

$$d_{j*} = \min_{0 \leq j \leq N-1} \{d_j\} \quad (3)$$

根据企业二次创业金融数据特征权重进行自适应分类,构建融资相关性决策模型,提高对数据的分类和统计决策能力。

### 1.2 模糊特征分析

构建企业二次创业金融数据分布的不规则空间聚类模型,采用相空间结构重组方法进行模糊特征重构,采用决策树模型进行统计融合分类<sup>[5]</sup>,得到企业二次创业金融数据的量化特征分布集为  $D, D =$

$\{S_{i,j}(t), T_{i,j}(t), U_{i,j}(t)\}$ ,其中  $S_{i,j}(t)$  表示特征权重的重复因素,  $T_{i,j}(t)$  表示融合分类的输出量因素,  $U_{i,j}(t)$  表示相似度(相关性)模型,对企业二次创业金融数据特征权重关联规则特征量进行量化回归分析,定义为

$$S_{i,j}(t) = \frac{p_{i,j}(t) - sp_{i,j}(t)}{p_{i,j}(t)} \quad (4)$$

$T_{i,j}(t)$  表示对企业二次创业金融数据特征权重检测的相关性特征分布集,计算表达为

$$T_{i,j}(t) = \frac{|p_{i,j}(t) - \Delta p(t)|}{p_{i,j}(t)} \quad (5)$$

提取企业二次创业金融数据的关联规则特征量,根据谱聚类结果,进行自适应筛选和优化决策,建立关联规则模型,输出为

$$U_{i,j}(t) = \exp[-b[z_i(t) - z_j(t)]^2] \quad (6)$$

式(6)中,  $p_{i,j}(t)$  为统计融合分类的互信息量;  $sp_{i,j}(t)$  为特征权重检测的分叉度重复量;  $\Delta p(t)$  为增益系数;  $z_i(t), z_j(t)$  表示为模糊度函数,由此建立企业二次创业金融数据的特征提取和大数据融合聚类模型,采用模糊特征分析方法,进行企业二次创业金融数据的模糊聚类处理,提高数据的分类挖掘和识别能力<sup>[6]</sup>。

## 2 企业二次创业金融数据分类优化

### 2.1 金融数据的关联规则特征提取

在构建企业二次创业金融数据优化分类模型中,采用融资决策模型,进行融合调度,提取关联规则特征集进行金融数据的自适应分类,分析企业二次创业金融数据的离散空间调度模型,采用模糊控制方法<sup>[7]</sup>,进行融分段样本检验,得到自适应加权权重为

$$\omega_j = (\omega_{0j}, \omega_{1j}, \dots, \omega_{(k-1)j})^T \quad (7)$$

建立企业二次创业金融数据的模糊聚类模型,进行特征分析,得到数据序列的统计特征序列为

$$\sum_{i=1}^c \mu_{ik} = 1, k = 1, 2, \dots, n \quad (8)$$

式(8)中,  $k$  为企业二次创业金融数据的聚类空间分布权重。采用离散序列调度方法, 构建企业二次创业金融数据的特征匹配模型, 根据多分类器融合结果进行信息分类, 实现企业二次创业金融数据的统计特征检测, 得到检测统计量为

$$V(a_1, \dots, a_m) = (a_i^{j-1})_{i,j=1}^m = \begin{pmatrix} 1 & a_1 & \dots & a_1^{m-1} \\ 1 & a_2 & \dots & a_2^{m-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_m & \dots & a_m^{m-1} \end{pmatrix} \quad (9)$$

$$A = (a_{i,j})_{i,j=1}^m = \begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,m} \end{pmatrix} \quad (10)$$

构建企业二次创业金融数据统计分布有限数据集模型, 得到关联特征为

$$E_{\text{total}} = (\alpha_1 + \alpha_2 d_1^{n_0}) + (\beta + \alpha_1 + \alpha_2 d_2^{n_0}) + \dots + (\beta + \alpha_1 + \alpha_2 d_p^{n_0}) \quad (11)$$

$$E_{\text{total}} = pa_1 + (p-1)b + a_2(d_1^{n_0} + d_2^{n_0} + \dots + d_p^{n_0}) \quad (12)$$

在分散子空间中进行企业二次创业金融数据实时检测, 实现关联规则特征提取<sup>[8]</sup>。

### 2.2 数据模糊 C 均值聚类和优化分类

提取企业二次创业金融数据的关联规则特征量, 采用逻辑回归分析方法进行数据分类的融合聚类处理, 采用关联规则分层调度方法, 进行特征权重聚类和可靠性检测<sup>[9]</sup>, 得到可靠性评估函数  $\hat{S}_w$  表述为

$$\hat{S}_w = \sum_{i=1}^c p_i \frac{1}{n_i} \sum_{k=1}^{n_i} [(\vec{X}_k^{(i)} - \vec{m}_i)(\vec{X}_k^{(i)} - \vec{m}_i)^T] \quad (13)$$

采用多队列调度方法, 建立企业二次创业金融数据的决策调度模型, 得到特征训练集  $s_i = \{x_j: d(x_j, y_i) \leq d(x_j, y_i)\}$ , 企业二次创业金融数据的显著性特征权重为

$$\text{Min } WH = \min\{w(cc), h(cc)\} \quad (14)$$

$$\text{Area\_Ratio} = \frac{\text{Area}(cc)}{\text{Area}(pic)} \quad (15)$$

建立模糊度核函数模型, 采用相空间结构重组方法, 得到  $N_{j^*}$  的几何邻域  $NE_{j^*}(t)$ , 企业二次创业金融数据统计融合分类的模糊聚类中心为

$$U = \{\mu_{ik} | i=1, 2, \dots, c; k=1, 2, \dots, n\} \quad (16)$$

在关联规则聚类下, 得到企业二次创业金融数据的模糊调度函数为

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m (d_{ik})^2 \quad (17)$$

采用空间网格聚类方法得到优化的企业二次创业金融数据特征权重聚类中心为

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \quad (18)$$

$$V_i = \frac{\sum_{k=1}^m (\mu_{ik})^m x_k}{\sum_{k=1}^n (\mu_{ik})^m} \quad (19)$$

式(18)(19)中,  $m$  为企业二次创业金融数据统计融合分类的适应度函数,  $(d_{ik})^2$  为样本  $x_k$  与特征聚类中心与样本  $V_i$  的测度距离。结合模糊 C 均值聚类方法, 得到企业二次创业金融数据主成分分布为

$$f(k) = \begin{cases} f(k-1) - \frac{1}{n}, & 1 \leq k < n \\ 1, & k = n \end{cases} \quad (20)$$

构建企业二次创业金融数据模糊聚类的回归分析模型, 根据特征分布进行线性结构重组, 得到企业二次创业金融数据的多元决策模型, 对其进行分布式检测和自适应聚类分析, 得到相关性概率密度特征为

$$D''_{i,j}(t_{n+1}) = \frac{D'_{i,j}(t_{n+1}) + f(n)D'_{i,j}(t_n)}{2} \quad (21)$$

在关联规则特征分布集中, 企业二次创业金融数据分类的互信息量为

$$I_{i,j}(t) = \frac{\sum D''_{i,k}(t)D''_{k,j}(t)}{\sum D''_{i,k}(t)} \quad (22)$$

建立企业二次创业金融数据的特征提取和大数据融合聚类分析模型, 实现特征权重分类<sup>[10]</sup>, 得到

企业二次创业金融数据统计分布序列特征矩阵, 满足:

$$A = \begin{pmatrix} t_{i_1,j_1} & t_{i_1,j_2} & \cdots & t_{i_1,j_k} \\ t_{i_2,j_1} & t_{i_2,j_2} & \cdots & t_{i_2,j_k} \\ \cdots & \cdots & \ddots & \cdots \\ t_{i_k,j_1} & t_{i_k,j_2} & \cdots & t_{i_k,j_k} \end{pmatrix} \neq 0 \quad (23)$$

根据上述分析,采用决策树模型,构建企业二次创业金融数据的空间聚类模型,结合统计特征分析方法,实现数据的均值聚类和自适应分类识别方法,实现对企业二次创业金融数据分类的自适应寻优和收敛性控制,提高数据的分类检测能力。

### 3 仿真实验与结果分析

为了验证本文方法在实现企业二次创业金融数据分类中的应用性能,进行仿真实验分析。采用 Matlab 和 C++进行数据分类的算法设计,对金融数据采样的样本数为 1 200,控制权重系数为 0.36,金融数据的模糊特征检测迭代次数为 800,采样周期  $T=0.86\text{ s}$ ,信息的扰动强度为  $SNR=0\sim-20\text{ dB}$ ,统计采样率为  $f_s=(10 * f_0)\text{ Hz}=10\text{ kHz}$ ,根据上述仿真环境和参数设定,进行企业二次创业金融数据统计融合分类,得到大数据集采样分布如图 1 所示。

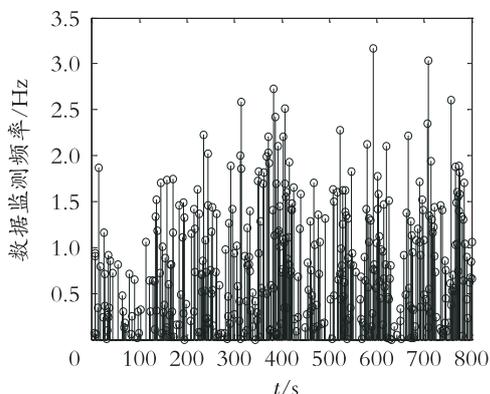


图 1 企业二次创业金融数据的大数据集采样频域分布  
Fig. 1 The sampling frequency distribution of the big data of the financial data for the secondary entrepreneurship of a firm

以图 1 的数据为研究对象,构建企业二次创业金融数据分布的不规则空间聚类模型,采用相空间结构重组方法进行模糊特征重构,得到企业二次创业金融数据的聚类中心分布如图 2 所示。

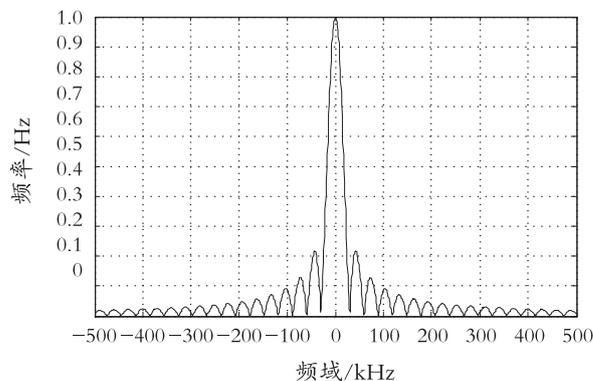


图 2 企业二次创业金融数据的聚类中心分布  
Fig. 2 The clustering center distribution of the financial data

分析图 2 得知,采用本文方法进行企业二次创业金融数据分类的空间聚集性较好,聚类中心检测的抗干扰能力较强。测试不同方法的企业二次创业金融数据的准确性,得到收敛性测试结果如图 3 所示。

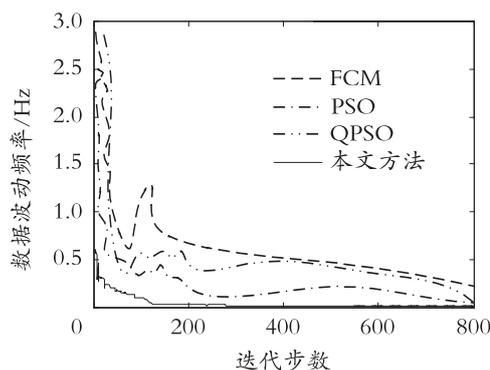


图 3 企业二次创业金融数据分类的收敛性结果  
Fig. 3 The convergence result of the financial data classification of the secondary entrepreneurship of a firm

分析图 3 得知,采用本文方法进行企业二次创业金融数据统计融合分类的准确性较高,进一步测试企业二次创业金融数据的误分率,得到对比结果见表 1。分析表 1 得知,本文方法进行企业二次创业金融数据统计分类的误分率较低。

表 1 误分率对比

Table 1 Comparison of error rate

迭代次数	本文方法	模糊 C 均值	PSO
100	0.117	0.254	0.156
200	0.087	0.164	0.134
300	0.032	0.132	0.074
400	0	0.065	0.032

## 4 结 语

结合大数据信息处理和数据融合方法,进行企业二次创业金融数据的优化聚类分析,提高企业二次创业金融数据的模糊聚类能力。提出基于逻辑回归的企业二次创业金融数据分类方法,采用自适应无监督学习方法进行统计融合处理,构建企业二次创业金融数据分布的不规则空间聚类模型,采用相空间结构重组方法进行模糊特征重构,采用逻辑回归分析方法进行融合聚类处理,结合模糊 C 均值聚类方法,实现对企业二次创业金融数据分类的自适应寻优和收敛性控制,实现分类优化。研究得知,采用该方法进行企业二次创业金融数据分类的准确性较高,误分率较低。

### 参考文献 (References):

- [1] NASCIMENTO F F D, SILVA W V M E. A Bayesian Model for Multiple Change Point to Extremes, with Application to Environmental and Financial Data [J]. *Journal of Applied Statistics*, 2016, 44(13):1—17
- [2] 张瑞,唐旭丽,王定峰,等. 基于知识关联的金融数据可视化分析[J]. *情报理论与实践*, 2018, 41(10):135—140
- ZHANG R, TANG X L, WANG D F, et al. Visualization Analysis of Financial Data Based on Knowledge Association[J]. *Information Studies: Theory & Application*, 2018, 41(10):135—140(in Chinese)
- [3] 罗明,黄海量. 一种基于语义标注特征的金融文本分类方法[J]. *计算机应用研究*, 2018, 35(8):2281—2284,2288
- LUO M, HUANG H L. New Approach of Financial Text Classification Based on Semantic Annotation Features[J]. *Application Research of Computers*, 2018, 35(8):2281—2284,2288(in Chinese)
- [4] 何培育. 基于互联网金融的大数据应用模式及价值研究[J]. *中国流通经济*, 2017, 31(5):39—46
- HE P Y. Research on the Application Model and Value of Big Data Based on Internet Finance[J]. *China Business and Market*, 2017, 31(5):39—46(in Chinese)
- [5] 陆莉莉,张永潘,谈海宇,等. 大数据分类挖掘算法及其概念漂移应用研究[J]. *计算机科学与探索*, 2016, 10(12):1683—1692
- LU L L, ZHANG Y P, TAN H Y, et al. Research on Classification Algorithm and Concept Drift Based on Big Data [J]. *Journal of Frontiers of Computer Science & Technology*, 2016, 10(12):1683—1692(in Chinese)
- [6] 张兆晨,冀俊忠. 基于卷积神经网络的 fMRI 数据分类方法[J]. *模式识别与人工智能*, 2017, 30(6):71—80
- ZHANG Z C, JI J Z. Classification Method of fMRI Data Based on Convolutional Neural Network [J]. *Pattern Recognition and Artificial Intelligence*, 2017, 30(6):71—80(in Chinese)
- [7] 刘东启,陈志坚,徐银,等. 面向不平衡数据分类的复合 SVM 算法研究[J]. *计算机应用研究*, 2018, 35(4):1023—1027
- LIU D Q, CHEN Z J, XU Y, et al. Hybrid SVM Algorithm Oriented to Classifying Imbalanced Datasets[J]. *Application Research of Computers*, 2018, 35(4):1023—1027(in Chinese)
- [8] 张宇,包研科,邵良杉,等. 面向分布式数据流大数据分类的多变量决策树[J]. *自动化学报*, 2018, 44(6):1115—1127
- ZHANG Y, BAO Y K, SHAO L S, et al. A Multivariate Decision Tree for Big Data Classification of Distributed Data Streams[J]. *Acta Automatica Sinica*, 2018, 44(6):

- 1115—1127(in Chinese)
- [9] 闫慈, 田翔华, 阿拉依·阿汗, 等. 基于 AdaBoost 法在代谢综合征不平衡数据分类中的应用[J]. 现代预防医学, 2017, 44(21):16—18,28
- YAN C, TIAN X H, ALAYI A, et al. Based on the Application of AdaBoost + Decision Tree for Metabolic Syndrome with Imbalanced Data[J]. Modern Preventive Medicine, 2017, 44(21):16—18,28(in Chinese)
- [10] HUANG Y C, TSAO C Y. Discovering Traders' Heterogeneous Behavior in High-Frequency Financial Data[J]. Computational Economics, 2018, 51(4):1—26

## Research on Classification Method of Enterprise Secondary Entrepreneurship Financial Data Based on Logical Regression

**LAI Hong-qing**

(Department of Business and Trade, School of Business Administration,  
Foshan Polytechnic College, Guangdong Foshan 528200, China)

**Abstract:** The optimal classification of enterprise secondary venture finance data can improve the statistical analysis ability of the data. The classification method of enterprise secondary venture finance data based on logical regression is put forward. The self-adaptive unsupervised learning method is used to carry on the statistical fusion processing of enterprise secondary venture finance data, and the irregular spatial clustering model of enterprise secondary venture finance data distribution is constructed. The fuzzy feature reconstruction of enterprise secondary venture finance data is carried out by using phase spatial structure reorganization method, the self-association rule feature quantity of enterprise secondary venture finance data is extracted, and the fusion clustering processing of enterprise secondary venture finance data classification is carried out by using logical regression analysis method. Combined with fuzzy C-means clustering method, the adaptive optimization and convergence control of enterprise secondary venture finance data classification are realized. To realize the classification and optimization of the secondary entrepreneurial financial data of enterprises, the simulation results show that the method has the advantages of high accuracy, good convergence and strong feature clustering ability.

**Key words:** logical regression; enterprise secondary entrepreneurship; financial data; classification

责任编辑:李翠薇

引用本文/Cite this paper:

赖红清. 基于逻辑回归的企业二次创业金融数据分类方法研究[J]. 重庆工商大学学报(自然科学版), 2021, 38(5):114—119

LAI H Q. Research on Classification Method of Enterprise Secondary Entrepreneurship Financial Data Based on Logical Regression [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2021, 38(5):114—119