

基于多维时间序列的灰色模糊信用评价研究^①

张洪祥^{1 2}, 毛志忠^{1 3}

(1. 东北大学信息科学与工程学院, 沈阳 110004; 2. 辽东学院信息技术学院, 丹东 118001;
3. 东北大学流程工业综合自动化教育部重点实验室, 沈阳 110004)

摘要: 传统信用评价技术多在孤立时间点上对受评目标数据进行评价分析, 但受评目标由于某些原因可能产生数据“突变”, 导致评价结果失真, 产生信用风险. 针对这一问题, 本文提出应用多维时间序列数据对受评样本进行信用评价. 该方法首先对多维时间序列数据使用灰色关联分析方法进行分割处理, 解决“维数灾难”带来的严重影响, 并将得到的灰色关联度值作为信用评判值; 再运用模糊聚类方法对信用评判值构成时间序列矩阵进行信用评价分析, 得到受评样本的“真实”信用等级. 通过实例验证, 该方法可以从时间序列的角度观察受评目标信用等级状态趋势及“波动”情况, 解决因为数据“突变”造成的评价结果失真问题, 具有良好的评价效果和实用价值.

关键词: 信用评价; 多维时间序列; 灰色关联分析; 模糊聚类分析

中图分类号: F830.59 **文献标识码:** A **文章编号:** 1007-9807(2011)01-0028-10

0 引言

由2007年美国次贷危机引起的全球性金融危机已经给全世界的金融市场造成了巨大经济损失, 全球股市、汇市、期货市场无一幸免. 这次金融危机根源就在于美国房地产市场上的次级按揭贷款危机. 次级按揭贷款是相对于给信用等级较高的按揭贷款而言的, 给缺乏足够的收入、没有还款能力或者其他负债较重的按揭贷款人提供房地产按揭贷款, 由于这批客户的信用等级较次, 所以就被称为次级按揭贷款. 正是这种大量的低等级信用客户不能够按时偿还贷款, 导致银行被动承担越来越多的亏损, 更使由次贷衍生的金融产品风险日益加巨, 最终导致全面的崩盘. 由此可见对信用风险的监管与控制, 对于当今经济金融领域有着多么重要的地位. 信用评价就是控制信用风险最重要、最有效的工具之一. 信用评价分析是指根据科学的指标体系, 采用严谨的分析方法, 运用简

明的文字符号, 对被受评企业或者个人履行经济责任的能力及其可信任程度进行客观公正的评价, 并确定其信用等级的一种经济活动. 无论对贷款企业还是个人信贷进行信用评价, 建立相应的信用风险评价模型, 都是为了能够使用评价模型预测违约风险可能性, 以便及早发现信用危机信号, 使金融机构可依据这种信号及时作出信贷风险决策. 因此建立信用评价系统, 可以提高金融机构在经济决策中审核质量, 对提高信贷资产的安全性并减少不良资产的产生有极其重要的现实意义.

信用评价技术诞生于20世纪30年代, 在经历几次重大的金融危机后, 美国与欧洲学者开始重视对信用风险的分析研究. 早期信用评价技术多使用统计方法对样本数据进行建模分析处理, 用统计方法对人群进行判别分类应用最早要追溯到1936年Fisher. Durand则是第一个使用判别技

^① 收稿日期: 2008-11-24; 修订日期: 2010-09-28.

基金项目: 国家自然科学基金资助项目(60774068); 973计划课题资助项目(2002CB312201).

作者简介: 张洪祥(1978-), 男, 辽宁丹东人, 博士生. Email: zhx780711@yahoo.com.cn

术对信贷客户进行好、坏分类的学者。1958 年 Bill Fair 和 Earl Isaac 为美国投资业建立了第一个信用评价系统。1966 年, Beaver 首次应用单变量判定分析来研究公司财务危机问题, 建立了财务危机预测模型^[1]。1977 年, Martin 采用一系列财务比率变量来预测公司破产及违约概率, 建立 Logistic 回归模型^[2]。1985 年, Messier 和 Hansen 使用专家系统应用到信用风险分析领域中^[3]。1996 年, Desai 等将模式神经网络(MNN)引入信用风险分析^[4]。Hashem 和 Blanc 等在 1998 年研究了商业银行控股结构的杂合智能系统^[5]。Reshmi 和 Malhotrad 2002 年利用神经模糊系统对消费贷款的 500 个训练样本进行信用评价, 分类准确率达到了 73.4% ~ 76%^[6]。

我国学者对信用评价研究起步较晚, 目前尚属于起步阶段。郝丽萍, 胡欣悦, 李丽利用 BP 网络模型对商业银行信贷风险进行分析和预测^[7]。杨保安, 季海利用 BP 网络算法对商业银行贷款风险进行预警研究^[8]。庞素琳和王燕鸣等在 2003 年通过建立多层感知器(MLP)信用评价模型, 对上市公司分别进行两类模式分类和三类模式分类, 分类准确率分别达 98.11% 和 79.17%。同年利用两类模式 MLP 信用评价模型对我国 2001 年公布的 13 家预亏公司进行预警分析, 结果表明预警准确率达到 100%^[9]。2004 年李晓峰, 徐玖平提出基于 Rough 集理论和 ANN(人工神经网络)方法建立了企业财务危机预警的 Rough-ANN 模型, 并就企业财务危机预警问题进行了实证分析^[10]。2005 年杨淑娥, 黄礼用 BP 神经网络对我国上市公司进行了财务预警, 取得了较好的效果^[11]。梁琪将主成分分析方法与 logistic 回归方法相结合应用于对企业经营管理预警, 效果较好^[12]。2007 年, 石晓军、任若恩、肖远文对边界 Logistic 违约率模型展开进一步研究, 通过抽样分布性质的研究, 证明了边界 Logistic 违约率模型优越性^[13]。同年程功, 张维, 熊熊研究在有噪音的信息环境下, 银行如何利用结构化模型来预测违约概率问题, 并利用结构化模型原理建立了违约概率模型^[14]。2009 年, 孙洁、李辉等通过多分类器混合组合模型来对企业财务困境进行预测, 实现了并联组合和串联组合的优势互补作用^[15]。同年杨海军, 太雷提出了基于模糊支持向量机模型方法对上市公

司财务困境进行预测, 精度也有所提高^[16]。王东静, 张祥建, 张景青利用理论模型对我国上市公司的短期债务比例与违约风险的关系进行实证研究^[17]。迄今为止, 国内外提出的信用评价方法已有很多, 总体上可归为几大类: 统计学方法、专家系统法、神经网络方法、支持向量机方法与杂合系统等^[18]。

上述的信用评价技术基本上多是在孤立时点上对受评目标数据进行评价分析, 评价的结果只针对当前时刻的受评样本数据, 然而这种数据处理方式往往会忽视了样本数据突变的问题。数据突变是指当某些受评目标由于特殊原因引起这一时点上的数据发生较大幅度增长或减小, 与平时稳态时的数据产生巨大的差异。这种现象在金融经济领域常常出现, 在这种极端情况下采集的数据来对受评主体进行信用评价, 就会产生较为严重错评现象, 评价结果往往缺乏真实性、客观性, 容易产生信用风险。如何能够真实有效地对受评目标进行信用评价, 则是本文主要解决的问题。

针对这个问题, 本文提出了基于多维时间序列数据的信用评价方法。该方法的基本思想在于对上市公司这样的受评样本, 由于其本身经济结构复杂性, 常常因为政策或者其它因素引起某段时刻的经营业绩发生较大的变化, 在财务数据上产生数据突变(巨幅增长或者减少), 特别在财务比率上尤为明显, 但从一个长期的经营周期来看, 每家上市公司都会呈现出一种较为稳定的经营业绩走势, 同样上市公司的信用状态也会在一段时间内呈现相对稳定的趋势, 这种现象可以从受评目标的多维时间序列数据(历史数据)挖掘出来。按照传统的信用评价技术, 如果在突变时点上进行评价, 就可能造成严重的错评结果, 而采用多维时间序列数据进行信用评价, 则可以从一段长期有效的采样时间点上观察受评目标的信用状态, 得到受评目标在这一时段内的信用波动表现, 找到受评样本信用状态的稳定趋势, 克服在孤立时间点上由于数据“突变”造成的评价效果失真问题, 得出较为客观真实的评价结果, 提高对信用风险的控制, 有效地规避风险。

本文将灰色关联分析与模糊聚类方法进行组合生成混杂灰色模糊聚类算法应用于多维时间序列数据下的信用评价分析。由于多维时间序列数

据信息量一般较大,同时维数增多容易引起“维数灾难”,直接处理非常困难,解决这一问题的有效方法就是对多维时间序列进行降维处理.本文首先对多维时间序列数据按照采样时间点进行分割处理,在每一个独立采样时间点上进行灰色关联分析,得到受评样本在各个时间采样点上的灰色关联度值并作为信用评判值,达到降维目的.将受评样本的信用评判值构成模糊聚类原始数据矩阵,通过将其转化为模糊相似矩阵,进行模糊聚类分析,得出受评样本的最终信用等级.该方法优点在于摒弃了以往孤立时间点上对受评样本进行的信用评价方式,从时间序列的角度出发分析受评样本的长期历史信用状态,通过信用评价时间序列曲线找出受评样本真实的信用趋势及波动性,分析出样本的真实信用级别状态,克服由于数据突变造成的孤立时间点上信用评价的不真实不客观的问题,大大提高了对受评样本信用风险的控制.

1 多维时间序列数据

定义 1 时间序列(time series)是一组变量在连续的时间上获得的一系列采样值.长度为 n 的 d 维时间序列^[19-20]

$$x_i(t); [i = 1, 2, \dots, d; t = 1, 2, \dots, n]$$

其中 $x_i(t)$ 是索引时间点 t 上第 i 维变量 x_i 的值.如果仅对单个变量进行采样,即 $d = 1$,那么所获得的时间序列称为一维时间序列.

相反,如果采样对象与一个时刻对应的往往是多个实数变量或者描述性的属性即 $d > 1$,而非简单的一维、实值变量,其相互之间的关系也不再局限于变量与时间之间的相关性,变量与变量之间也会具有一定的联系,以至于往往需要用图而不仅仅是多维向量来表示变量之间的关系,那么所获得的时间序列则称为高维时间序列.一般地,一个长度为 n 的 $d \times n$ 维时间序列可以用一个

$$[x_{ij}]^{d \times n} \text{ 矩阵表示 } \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & & \vdots \\ x_{d,1} & \dots & x_{d,n} \end{bmatrix}$$

矩阵中项 (i, t) 为第 i 维变量 x_i 在时间点 t 上的采样值.

由于多维时间序列是典型的高维数据,为了避免由于维数过多引起的所谓维数灾难,通常要对多维时间序列进行转换即降维处理.本文通过灰色关联分析对多维时间序列数据根据采样时间点进行分割处理.

2 信用评价等级与指标体系的建立

2.1 信用评价等级的划分

信用风险管理中最重要的问题是,金融机构包含多少信用等级.在传统银行评级系统中,信用的序数刻度非常重要(见表 1).

表 1 信用评价等级标准

Table 1 Credit evaluation grade standard

等 级	解 释
1	违约后不能收回欠款
2	违约后收回欠款的机会很小
3	违约后一般有机会收回欠款
4	违约后有很大机会收回欠款
5	接近违约水平
6	信用质量差
7	平均信用质量
8	信用质量好
9	优秀的信用质量
10	最高信用质量

这种序数刻度的等级划分运用的历史虽然较长,然而大多数银行金融机构常常只将评价对象归入信用等级中的 6、7、8 项,导致实际上只有这三个等级.本文风险等级划分也遵从这三个评价序数刻度,即:信用质量差、平均信用质量(中)、信用质量好三类^[21].

2.2 指标体系的建立

信用风险评价指标体系的建立必须要遵从我国的财会准则和相关法律,只有根据我国财务数据的特点,从中提取特征指标建立相应的信用评价模型,才能对我国企业的信用风险评价做出更为准确的预测.2000 年陈元燮对建立信用评级指标体系进行了比较深入的理论研究^[22-23].

本文参照前人对信用评价指标体系的研究,依据对其设置的原则:1) 全面性、2) 科学性、3) 针对性、4) 公正性、5) 合法性、6) 可操作性.从受

评样本的财务报表中提取相关能够反应其主要财务特征的财务比率构建信用评价指标体系(见表 2)。

表 2 信用评价指标体系
Table 2 Credit evaluation index system

信用评价目标层	一级指标	二级指标
综合信用评判值	1、风险抵偿能力评价	1. 资产负债率 2. 现金流动负债比 3. 股东权益比 4. 流动负债率 5. 负债权益比
	2、经营效率能力评价	1. 每股净资产 2. 每股资本公积金 3. 存货周转率 4. 应收账款周转率 5. 总资产周转率
	3、盈利水平能力评价	1. 每股收益 2. 资产利润率 3. 净利润率 4. 净资产收益率 5. 税前利润增长率
	4、未来成长性能力评价	1. 主营收入增长率 2. 净利润增长率 3. 总资产增长率 4. 每股资本公积金 5. 主营利润增长率

3 灰色模糊聚类方法模型建立

3.1 原始数据标准化

在信用指标体系中的各个指标度量单位是不同的,为了能够将指标参与评价计算,需要对指标进行规范化处理,通过函数变换将其数值映射到某个数值区间. 数据标准化模型如下:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{S_i} \quad (1)$$

其中 x_{ij} 为第 j 个样本第 i 项变量的原始数据

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}; S_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \quad (2)$$

这样得到的标准化数据 x'_{ij} 还不一定在 $[0, 1]$ 闭区间,为了把标准化数据压缩在 $[0, 1]$,再采用极值标准化公式

$$x''_{ij} = \frac{x'_{ij} - x'_{k \min}}{x'_{k \max} - x'_{k \min}} \quad (3)$$

这里 $x'_{k \max}$ 和 $x'_{k \min}$ 分别指 $x'_{1k}, x'_{2k}, \dots, x'_{nk}$ 中的最大值和最小值.

3.2 灰色关联分析

本文对多维时间序列数据进行分割处理,对每个采样时间点上的多维时间序列数据运用灰色关联分析方法进行的降维处理^[24-26].

3.2.1 确定参考序列和比较序列

所谓“参考序列”,常记为 x_{0j} ,它一般由最优单目标优化值构成. 参考序列 x_{0j} 可表示为

$$\begin{cases} X_{0j} = (x_{0j}(1), x_{0j}(2), \dots, x_{0j}(n)) \quad j = 1, 2, \dots, J \\ x_{0j}(k) \in x_{0j}, k \in K, K = 1, 2, \dots, n, n \geq 2 \end{cases}$$

其中,对多目标优化来说 k 为目标指标.

关联分析中与参考序列做关联程度比较的“子数列”称之为“比较序列”,又称为“系统相关因素行为序列”,记为 x_i . 比较序列可表示为

$$\begin{cases} X = (x_i(1), x_i(2), \dots, x_i(n)) \quad i = 1, 2, \dots, m \\ x_i(k) \in x_i, k \in K, K = 1, 2, \dots, n, n \geq 2 \end{cases}$$

3.2.2 计算灰色关联系数

$X_{0j} = (x_{0j}(1), x_{0j}(2), \dots, x_{0j}(n))$, $j = 1, 2, \dots, J$ 作为基准指标变量,若有非负实数 $\xi_{0i}(k)$ 为 X 上在一定环境下 $x_0(k)$ 与 $x_i(k)$ 的比较测度, $|x_0(k) - x_i(k)|$ 越小, $\xi_{0i}(k)$ 越大时,称 $\xi_{0i}(k)$ 为 x_i 对 x_0 在 k 点的灰色关联系数.

灰色关联系数为

$$\xi_{0i}(j) =$$

$$\frac{\min_{j \in M} \min_{k \in K} |x_0(k) - x_i(k)| + \rho \max_{i \in M} \max_{k \in K} |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_{i \in M} \max_{k \in K} |x_0(k) - x_i(k)|} \quad (4)$$

其中, $|x_0(k) - x_i(k)|$ 是距离,而 $\min_{j \in M} \min_{k \in K} |x_0(k) - x_i(k)|$, $\max_{i \in M} \max_{k \in K} |x_0(k) - x_i(k)|$ 是 x_i 与 x_0 的比较环境,也是 $x_i(k)$ 的领域,它含有点集拓扑信息. 常数 ρ 称为分辨系数 $\rho \in [0, 1]$. 它的作用在于调整比较环境的大小. 当 $\rho = 0$ 时,环境消失;当 $\rho = 1$ 时环境被“原封不动”地保持着. 因此,一般取 $\rho = 0.5$.

3.2.3 计算灰色关联度

两个系统或者两个因素间关联性大小的度量,称为关联度. 关联度描述了系统发展过程中,因素间相对变化的情况,也就是变化大小、方向与速度等的相对性. 灰色关联度公式为

$$w(x_0, x_i) = \sum_{k=1}^n \lambda(k) \xi_{0i}(k) \quad (5)$$

其中 $\sum_{k=1}^n \lambda(k) = 1, 0 \leq \lambda(k) \leq 1$ (指标权重).

对于高维数据而言参考序列和比较序列都可能不止一个,则各比较序列对各参考序列的灰色关联度则构成灰色关联矩阵,并可以作为相关分析的基础.若 $w_{ij}(i = 1, 2, \dots, n, j = 1, 2, \dots, m)$,便可得到灰色关联矩阵:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nm} \end{bmatrix} \text{ 或}$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix}$$

3.3 模糊聚类

模糊聚类是属于基于目标函数的聚类方法.它适用于大数据量的情况,能够满足实时性要求较高的场合,因此在实际中应用广泛.特别是当样本较大时,往往模糊相似矩阵也过大,但是由于模糊聚类本身是通过将模糊相似矩阵进行平方自合成运算,并通过截取合适水平截值进行分类,所以当矩阵出现病态现象时也不会影响分类的结果,这也是模糊聚类优势^[27-29].

3.3.1 模型相似矩阵

对得到的灰色关联度矩阵 $W_{[i,j]}$ 要将其转化为模糊相似矩阵 \tilde{R} .

本文采用数量积法来计算

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{M} \sum_{k=1}^m x'_{ik} x'_{jk} & i \neq j \end{cases} \quad (6)$$

其中 $M \geq \max_{i \neq j} \left(\sum_{i=1}^m x'_{ik} x'_{jk} \right)$;

$$\tilde{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix}$$

得到的模糊相似矩阵通常只满足自反性 ($r_{(i,i)} = 1$) 和对称性 ($r_{(i,j)} = r_{(j,i)}$).

3.3.2 传递闭包及聚类

模糊相似矩阵并不能直接进行分类,这是因为它本身并不具备传递性.因此需要将模糊相似矩阵改造成等价矩阵,只需要求相似矩阵的传递闭包^[34].

本文利用扎德算子对模糊相似矩阵进行平方自合成运算,求得

$$\tilde{R}^* = \tilde{R} \circ \tilde{R} \circ \dots \circ \tilde{R} \quad (7)$$

\tilde{R}^* 称为包含 \tilde{R} 的最小模糊等价矩阵.其中, $\tilde{R}^K = (r_{ij}^*)$; 称为 \tilde{R} 的传递闭包,其元素 r_{ij}^* 定义为

$$r_{ij}^* = \begin{cases} 1, & \text{当 } r_{ij} \geq \lambda \\ 0, & \text{当 } r_{ij} < \lambda \end{cases}$$

λ 可取 $[0, 1]$ 内的任意实数.当 $r_{ij} \leq \lambda$, 样本 x_i 与 x_j 可并为同一类.所以,适当的水平截值 λ 时,可以进行分类.

4 实例验证

4.1 数据选取

本文选取深、沪两市某板块 9 家上市公司作为受评样本,其中对于 9 家受评上市公司的选取,是根据其经营业绩优劣以及在行业板块中的所处地位(上游 3 家、中游 3 家、下游 3 家)次序选取.同时对样本数据选择并不受样本分布的限制,也不刻意剔除奇异数据或删除不合格数据.这是本文采用多维时间序列数据分析与传统方法上样本数据选取方法的不同之处.

本文以 2003 年 6 月到 2008 年 3 月季报为时间采样点,指标变量依据信用评价指标体系选取(指标体系见表 1),数据分别从各季财务报表中提取的财务比率作为信评原始数据.构成采集时间点为 $T = 20$ 样本为 $N = 9$,变量长度为 $i = 20$ 的多维时间序列数据库 $[x_{ij}(N)]^{20 \times 20}$.

4.2 多维时间序列数据的分割处理

本文对采集的多维时间序列数据按照采样时间点进行分割处理,将多维时间序列数据分割成 20 个数据库,分别对每个采样时点上的受评样本数据进行灰色关联分析,得出每个样本的信用评判值(灰色关联度).对 20 个采样时间点上的数据进行仿真,便可得到受评样本的信用状态趋势以

及其波动情况(见图 1、图 2、图 3)。

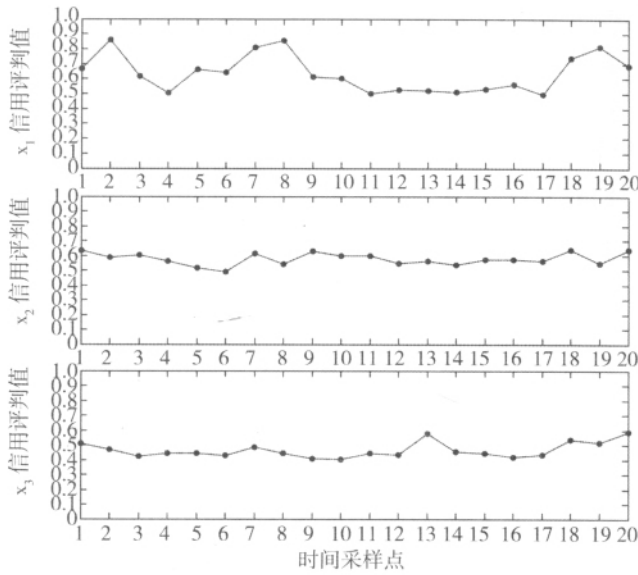


图 1 信用评价曲线(x_1, x_2, x_3)

Fig.1 Credit evaluation curve(x_1, x_2, x_3)

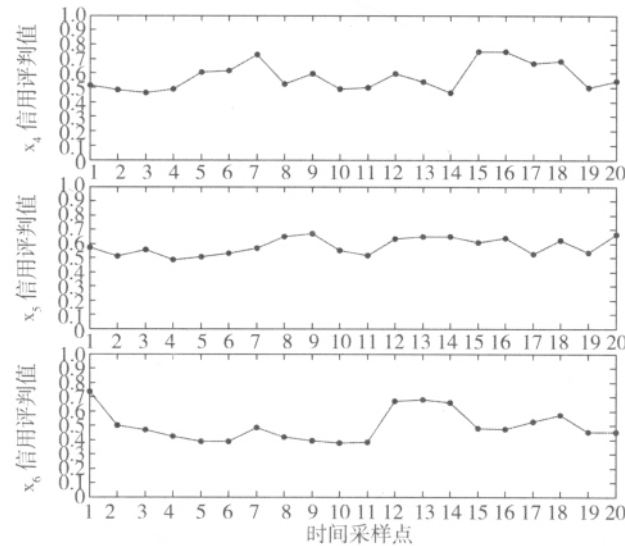


图 2 信用评价曲线(x_4, x_5, x_6)

Fig.2 Credit evaluation curve(x_4, x_5, x_6)

本文使用 MATLAB7.0 环境进行编程、仿真。

4.3 模糊聚类分析

通过上述对多维时间序列的预处理,可以分别得到受评样本在各个采样时间点上的信用评判值。接下来就可以进行最终的信用评价分析。将受评样本在各个时点上的信用评判值构成的一维时间序列向量,利用这些数据组成模糊聚类分析的原始数据矩阵,再通过数量积法转化为模糊相似矩阵。

$$\tilde{R} = \begin{bmatrix} 1.00 & 0.91 & 0.74 & 0.91 & 0.92 & 0.75 & 0.97 & 0.89 & 0.79 \\ 0.91 & 1.00 & 0.67 & 0.83 & 0.84 & 0.68 & 0.88 & 0.82 & 0.72 \\ 0.74 & 0.67 & 1.00 & 0.66 & 0.67 & 0.55 & 0.71 & 0.65 & 0.58 \\ 0.91 & 0.83 & 0.66 & 1.00 & 0.83 & 0.68 & 0.87 & 0.81 & 0.71 \\ 0.92 & 0.84 & 0.67 & 0.83 & 1.00 & 0.69 & 0.89 & 0.82 & 0.72 \\ 0.75 & 0.68 & 0.55 & 0.68 & 0.69 & 1.00 & 0.73 & 0.66 & 0.59 \\ 0.97 & 0.88 & 0.71 & 0.87 & 0.89 & 0.73 & 1.00 & 0.87 & 0.76 \\ 0.89 & 0.82 & 0.65 & 0.81 & 0.82 & 0.66 & 0.87 & 1.00 & 0.71 \\ 0.79 & 0.72 & 0.58 & 0.71 & 0.72 & 0.59 & 0.76 & 0.71 & 1.00 \end{bmatrix}$$

利用扎德算子对模糊相似矩阵进行平方自合成运算,求出传递闭包

$$\tilde{R}^* = \begin{bmatrix} 1.00 & 0.91 & 0.74 & 0.91 & 0.92 & 0.75 & 0.97 & 0.89 & 0.79 \\ 0.91 & 1.00 & 0.74 & 0.91 & 0.91 & 0.75 & 0.91 & 0.89 & 0.79 \\ 0.74 & 0.74 & 1.00 & 0.74 & 0.74 & 0.74 & 0.74 & 0.74 & 0.74 \\ 0.91 & 0.91 & 0.74 & 1.00 & 0.91 & 0.75 & 0.91 & 0.89 & 0.79 \\ 0.92 & 0.91 & 0.74 & 0.91 & 1.00 & 0.75 & 0.92 & 0.89 & 0.79 \\ 0.75 & 0.75 & 0.74 & 0.75 & 0.75 & 1.00 & 0.75 & 0.75 & 0.75 \\ 0.97 & 0.91 & 0.74 & 0.91 & 0.92 & 0.75 & 1.00 & 0.89 & 0.79 \\ 0.89 & 0.89 & 0.74 & 0.89 & 0.89 & 0.75 & 0.89 & 1.00 & 0.79 \\ 0.79 & 0.79 & 0.74 & 0.79 & 0.79 & 0.75 & 0.79 & 0.79 & 1.00 \end{bmatrix}$$

最终当 $\lambda = 0.91$ 时,将受评样本分为 $\{x_1, x_2, x_5, x_7\}$ 信用好、 $\{x_4, x_8\}$ 信用中、 $\{x_3, x_6, x_9\}$ 信用差三类。

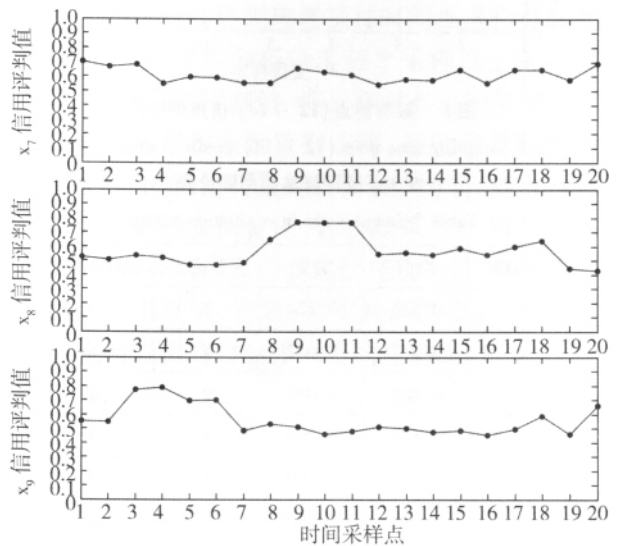


图 3 信用评价曲线(x_7, x_8, x_9)

Fig.3 Credit evaluation curve(x_7, x_8, x_9)

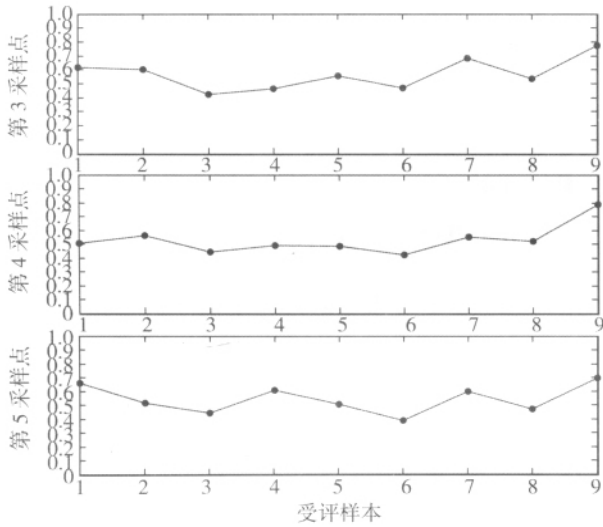


图4 采样时点(3~5)信用评判值

Fig.4 Sampling time point(3~5) credit evaluation value

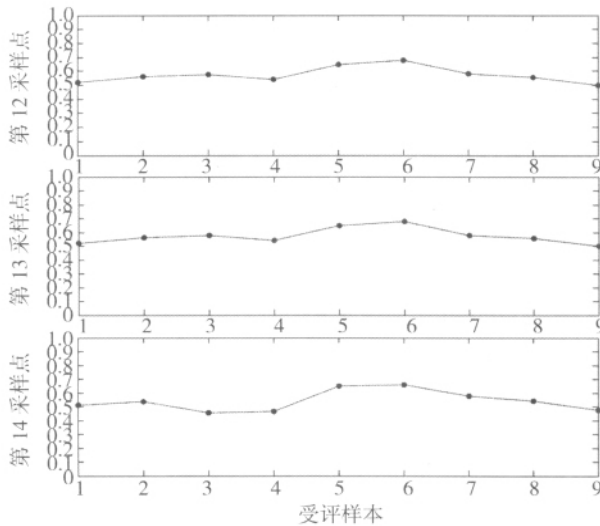


图5 采样时点(12~14)信用评判值

Fig.5 Sampling time point(12~14) credit evaluation value

表3 受评样本信用评价表

Table 3 Samples credit evaluation results

受评样本	均值	方差	排名	信用等级
x_1	0.636	0.124	1	I
x_2	0.577	0.042	4	II
x_3	0.464	0.053	9	III
x_4	0.575	0.096	5	I
x_5	0.581	0.060	3	I
x_6	0.475	0.092	8	III
x_7	0.613	0.051	2	I
x_8	0.567	0.104	6	II
x_9	0.500	0.087	7	III

4.4 结果分析

从(图1~图3)上可以观察到受评样本在采样时点上信用状态趋势及其波动性. 其中 x_1, x_4, x_6, x_8, x_9 的曲线波动幅度较大; x_2, x_3, x_5, x_7 的曲线振幅较小, 趋势较为平稳. 通过计算可以得到它们的方差值(见图3). 通过(表4)可以看到受评样本在每个时间采样点上的信用评判值, 结合(图1、图2、图3)可以看到某些受评样本由于数据突变导致财务指标大幅增长或减少而引起的评价失真的问题. 本文以信用级别差的两个样本予以说明: x_9 信用级别差, 最高值为 0.787 ($t = 4$), 最低值为 0.452 ($t = 16$), 均值为 0.500, 方差为 0.087, 该样本在第3~6采样点上由于财务数据发生巨大增幅, 导致在这4个采样点的信用评判值排名第一(图4), 然而在随后各个采样点上该样本又回到排名尾部, 通过(图3) x_9 的信用曲线可以看到该样本方差值较大. x_6 信用级别差, 最高值为 0.733 ($t = 1$), 最低值为 0.376 ($t = 10$), 均值为 0.475, 方差为 0.092, 该样本在第1和12~14采样点上由于财务数据发生巨大增幅, 导致在这4个采样点的信用评判值排名靠前(图5), 然而在其它采样点上该样本又回到排名尾部, 通过(图2) x_6 的信用曲线可以看到, 该样本方差较大, 说明波动性大. 通过对受评样本 x_6, x_9 在信用评价曲线以及采样时间点上的分析, 可以看到受评样本如果产生财务数据突变, 就可能会使在孤立时间点上进行的信用评价结果产生错误, 增大发生信用风险的概率.

从最终的信用评价结果中, 可以看到排名第1的 x_1 信用级别为好, 信用评判均值 0.636, 采样时段股价最高为 73.50 元, 最后采样时刻股价为 40.49 元; 排名第2的 x_7 信用级别为好, 信用评判均值 0.613, 采样时段股价最高为 102.89 元, 最后采样时刻股价为 51.72 元; 排名第3的 x_5 信用级别为好, 信用评判均值 0.581, 采样时段股价最高为 40.78 元, 最后采样时刻股价为 23.25 元; 排名第4的 x_2 信用级别为中, 均值 0.577, 以采样时段股价最高为 32.20 元, 最后采样时刻股价为 20.99; 排名第5的 x_4 信用级别为中, 均值 0.575, 采样时段股价最高为 20.35 元, 最后采样时刻股价为 12.19 元; 排名第6的 x_8 信用级别为中, 均值 0.567, 采样时段股价最高为 22.98 元, 最后采样时刻股价为 15.00 元; x_9 信用级别为差, 排名第7,

信用评判均值 0.500, 采样时段股价最高为 27.06 元, 最后采样时刻股价为 18.39 元; x_6 信用级别为差, 排名第 8, 均值 0.475, 采样时段股价最高为 14.29 元, 最后采样时刻股价为 9.78 元; x_3 信用级

别为差, 排名第 9, 均值 0.464, 采样时段股价最高为 19.75 元, 最后采样时刻股价为 12.37 元. 通过与受评样本的股价表现、经营业绩以及行业综合实力地位对比, 该方法的信用评价结果真实、客观.

表 4 受评样本时间序列评判值表

Table 4 Samples time series evaluation values

受评样本 / 采样时间	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	0.668	0.633	0.508	0.515	0.572	0.733	0.701	0.527	0.551
2	0.860	0.590	0.468	0.483	0.512	0.502	0.667	0.506	0.546
3	0.616	0.602	0.422	0.463	0.555	0.467	0.681	0.533	0.771
4	0.507	0.564	0.446	0.490	0.486	0.420	0.551	0.522	0.787
5	0.661	0.514	0.446	0.606	0.507	0.386	0.595	0.470	0.695
6	0.641	0.491	0.426	0.615	0.529	0.387	0.591	0.466	0.699
7	0.809	0.612	0.483	0.730	0.567	0.486	0.554	0.486	0.483
8	0.856	0.543	0.444	0.526	0.647	0.419	0.558	0.649	0.528
9	0.612	0.628	0.409	0.595	0.667	0.390	0.655	0.772	0.510
10	0.600	0.560	0.403	0.489	0.552	0.376	0.630	0.767	0.457
11	0.501	0.595	0.445	0.497	0.514	0.380	0.610	0.768	0.478
12	0.526	0.544	0.435	0.597	0.633	0.669	0.544	0.544	0.509
13	0.521	0.562	0.575	0.541	0.647	0.677	0.579	0.556	0.498
14	0.511	0.534	0.453	0.462	0.647	0.661	0.576	0.541	0.474
15	0.531	0.574	0.444	0.748	0.605	0.477	0.645	0.587 2	0.486
16	0.560	0.570	0.415	0.749	0.638	0.476	0.554	0.539	0.452
17	0.495	0.564	0.434	0.665	0.525	0.524	0.648	0.596	0.493
18	0.738	0.64	0.538	0.680	0.622	0.573	0.644	0.637	0.587
19	0.815	0.548	0.516	0.498	0.537	0.453	0.577	0.444	0.456
20	0.690	0.640	0.586	0.545	0.665	0.454	0.690	0.429	0.657

5 结束语

针对孤立时点上进行信用评价时, 由于某些特殊原因造成受评样本采集的数据产生突变(巨幅增长或减少), 从而使评价结果失真的问题, 本文提出基于多维时间序列数据进行信用评价的方法, 该方法优点在于从时间序列角度分析受评样本的历史信用状态, 通过仿真信用评价曲线研究受评样本真实的信用趋势及波动情况, 综合受评样本在各采样时点上的信用评判值进行最终的信

用评价分析, 得到真实信用级别. 本文使用灰色模糊聚类算法对多维时间序列数据进行分析处理, 首先对多维时间序列按照采样时间点进行分割处理, 利用灰色关联分析处理每个采样点上的多维数据, 得到信用评判值. 将采样时点上的信用评判值构成模糊聚类原始数据矩阵, 通过将其转化为模糊相似矩阵, 进行最终模糊聚类分析. 通过对样本上市公司的实例验证, 评价结果与上市公司的股价、经营业绩以及行业地位完全一致, 得到的信用评价结果真实、客观, 具有较高的说服力, 证明该方法可以提高对信用风险的控制能力.

参考文献:

- [1] Beaver W H. Financial ratios as predictors of failure [J]. Empirical Research in Accounting: Selected Studies, 1966, 4: 71 - 111.
- [2] Earky M D. Warning of bank failure: A logit regression approach [J]. Journal of Banking and Finance, 1977, 2: 249 - 276.
- [3] Messier W F, Hansen J V. Inducing rules for expert system development an example using default and bankruptcy data [J]. Management Science, 1985, 9: 253 - 266.
- [4] Desai V S, et al. A comparison of neural networks and linear scoring models in the credit union environment [J]. European

- Journal of Operational Research ,1996 ,95: 24 -37.
- [5] Hashemi R R , Le Blanc L A , Rucks C T , et al. A hybrid intelligent system for predicting bank holding structure [J]. European Journal of Operational Research ,1998 ,109: 211 -390.
- [6] Malhotra R , Malhotra D K. Differentiating between good credits and bad credit using neuro-fuzzy systems [J]. European Journal of Operational Research ,2002 ,136: 190 -211.
- [7] 郝丽萍,胡欣悦,李 丽. 商业银行信贷风险分析的人工神经网络模型研究 [J]. 系统工程理论与实践,2001 ,21(5): 66 -69.
Hao Liping , Hu Xinyue , Li Li. Research on the artificial neural network model for the credit risk analysis of commercial banks [J]. Systems Engineering: Theory & Practice ,2001 ,21(5): 66 -69. (in Chinese)
- [8] 杨保安,季 海. 基于人工神经网络的商业银行贷款风险进行预警研究 [J]. 系统工程理论与实践,2001 ,21(5): 70 -74.
Yang Baoan , Ji Hai. A study of commercial bank loans risk early warning based on BP neural network [J]. Systems Engineering Theory & Practice ,2001 ,21(5): 70 -74(in Chinese)
- [9] 庞素琳,王燕鸣,罗育中. 多层感知器信用评价模型及预警研究 [J]. 数学的实践与认识,2003 ,33(9): 55 -62.
Pang Sulin , Wang Yanming , Luo Yuzhong. Study on multilayer perceptron credit scoring model [J]. Mathematics in Practice and Theory ,2003 ,33(9): 55 -62. (in Chinese)
- [10] 李晓峰,徐玖平. 企业财务危机预警 Rough: ANN 模型的建立及其应用 [J]. 系统工程理论与实践,2004 ,24(10): 8 -14.
Li Xiaofeng , Xu Jiuping. The establishment of rough-ANN model for pre-warning of enterprise financial crisis and its application [J]. Systems Engineering: Theory & Practice ,2004 ,24(10): 8 -14. (in Chinese)
- [11] 杨淑娥,黄 礼. 基于 BP 神经网络的上市公司财务预警模型 [J]. 系统工程理论与实践,2005 ,25(1): 12 -19
Yang Shu-e; Huang Li. Financial crisis warning model based on BP neural network [J]. Systems Engineering: Theory & Practice ,2005 ,25(1): 12 -19(in Chinese)
- [12] 梁 琪. 企业经营管理预警: 主成分分析在 logistic 回归方法的应用 [J]. 管理工程学报,2005 ,1: 100 -103.
Liang Qi. Enterprise running and management warning: Applications of principal components logistic model [J]. Journal of Management Engineering ,2005 ,1: 100 -103. (in Chinese)
- [13] 石晓军,任若恩,肖远文. 边界 Logistic 违约率模型及实证研究 [J]. 管理科学学报,2007 ,3: 44 -51
Shi Xiaojun , Ren Ruoen , Xiao Yuanwen. Bounded logistic default risk model and empirical evidences from China [J]. Journal of Management Sciences in China ,2007 ,3: 44 -51. (in Chinese)
- [14] 程 功,张 维,熊 熊. 信息噪音结构化模型与银行违约概率度量 [J]. 管理科学学报,2007 ,10(4): 38 -48.
Cheng Gong , Zhang Wei , Xiong Xiong. Noisy information structural model and bank evaluation of default probability [J]. Journal of Management Sciences in China ,2007 ,10(4): 38 -48. (in Chinese)
- [15] 孙 洁,李 辉,Zhang Meng. 企业财务困境的多分类器混合组合预测 [J]. 系统工程理论与实践,2009 ,2: 78 -86
Sun Jie , Li Hui , Zhang Meng. Multiple classifiers hybrid combination for companies' financial distress prediction [J]. Systems Engineering: Theory & Practice ,2009 ,2: 78 -86. (in Chinese)
- [16] 杨海军,太 雷. 基于模糊支持向量机的上市公司财务困境预测 [J]. 管理科学学报,2009 ,3: 102 -110.
Yang Haijun , Tai Lei. Predicting financial distress of listed corporations based on fuzzy support vector machine [J]. Journal of Management Sciences in China ,2009 ,3: 102 -110. (in Chinese)
- [17] 王东静,张祥建,张景青. 公司债券期限结构与违约风险 [J]. 管理科学学报,2009 ,2: 77 -82.
Wang Dongjing , Zhang Xiangjian , Zhang Jingqing. Corporate debt maturity structure and default risk [J]. Journal of Management Sciences in China ,2009 ,2: 77 -82. (in Chinese)
- [18] 庞素琳. 信用评价与股市预测模型研究及应用 [M]. 北京: 科学出版社,2005: 253.
Pang Sulin. Credit Rating and The Stock Market Forecasting Model and its Application [M]. Beijing: Science Press ,2005: 253. (in Chinese)
- [19] 杨叔子,吴 雅. 时间序列分析的工程应用. 武昌: 华中理工大学出版社,1991.
Yang Shuzi , Wu Ya. Engineering Application of Time Series Analysis [M]. Wuchang: Huazhong University of Science & Technology Press ,1991. (in Chinese)
- [20] 欧阳为民,蔡庆生. 数据库中的时态数据发掘研究 [J]. 计算机科学,1998 ,25(4): 60 -63.
Ouyang Weimin , Cai Qingsheng. Research on temporal data mining in databases [J]. Computer Science ,1998 ,25(4):

- 60-63. (in Chinese)
- [21]唐纳德·范·戴维特, 今井贤志. 信用风险模型与巴塞尔协议[M]. 北京: 中国人民大学出版社, 2005: 62-65
Danald Van Devener, Kenji Lmni. Credit Risk Models and the Basel Accords [M]. Beijing: China Renmin University Press, 2005. 62-65. (in Chinese)
- [22]陈元燮. 建立信用评级指标体系的几个理论问题[J]. 财经问题研究 2000, (8): 3-8
Chen Yuanxi. Several theoretical issues about the establishment of credit evaluation indexes system [J]. Research on Financial and Economic Issues, 2000 (8): 3-8. (in Chinese)
- [23]程 功, 任宇航. 公司资本结构对信用风险的影响[J]. 北京理工大学学报(社会科学版), 2007, 19(1): 73-76.
Cheng Gong, Ren Yuhang. The effect of company capital structure on the credit risk of bank [J]. Journal of Beijing Institute of Technology (Social Sciences Edition), 2007, 19(1): 73-76. (in Chinese)
- [24]Kung Chang-Yung, Wen Kun-Li. Applying grey relational analysis and grey decision-making to evaluate the relationship between company attributes and its financial performance: A case study of venture capital enterprises in Taiwan [J]. Decision Support Systems, 2007, 43(3): 842-852.
- [25]KuoYiyo, Yang Taho, Huang Guan-wei. The use of grey relational analysis in solving multiple attribute decision-making problems [J]. Computers & Industrial Engineering, 2008, 55(1): 80-93.
- [26]刘思峰, 党耀国, 方志耕. 灰色系统理论及其应用[M]. 北京. 科学出版社, 2004: 332
Liu Sifeng, Dang Yaoguo, Fang Zhigeng. Gray System Theory and Its Application [M]. Beijing. Science Press, 2004: 332. (in Chinese)
- [27]Luo Y, Pang S, Qiu S. Fuzzy Cluster in Credit Scoring. Proceedings of 2003 International Conference on Machine Learning and Cybernetics, China, 2003, 5(5): 2731-2736.
- [28]Pang S, Hu B, Bai Y, Application of fuzzy cluster analysis in credit scoring [J]. Journal of Systems Science and information, 2005, 38(1): 1027-1032.
- [29]诸克军, 成金华, 郭海湘. 模糊软分类中最佳聚类数的确定 [J]. 管理科学学报, 2005, 8(3): 8-14.
Zhu Kejun, Cheng Jinhua, Guo Haixiang. Optimal number of clusters in fuzzy soft clustering [J]. Journal of Management Sciences in China, 2005, 8(3): 8-14. (in Chinese)

Research of multidimensional time series credit evaluation based on gray-fuzz analysis model

ZHANG Hong-xiang^{1,2}, MAO Zhi-zhong^{1,3}

1. School of Information Science & Engineering, Northeastern University, Shenyang 110004, China;
2. College of Information Technology, Eastern Liaoning University, Dandong 118001, China;
3. Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Ministry of Education, Shenyang 110004, China

Abstract: Traditional credit evaluation technique only evaluates and analyzes the objective data at the isolated time points, but the objective may generate data variation due to some reasons. Therefore, this may lead to infidelity of the evaluation results and serious credit risk. Aimed at this problem, the multidimensional time series data is used to credit evaluation in this paper. First, the multidimensional time series data is dealt with separately by grey relational analysis, in order to solve the serious effects caused by dimension disaster. At the same time, the credit evaluation value is obtained. Then the time series matrix is used to fuzzy clustering analysis. The real credit evaluation grade is got. The results of our experiment showed this method can observe the credit grade fluctuation of the credit evaluated in view of the time series. Thus, the infidelity of the evaluation results caused by data variation problem is solved, and the proposed method has accurate evaluation effect and better performance in practical application.

Key words: credit evaluation; multidimensional time series; grey relational analysis; fuzzy cluster analysis