

# 基于马氏距离判别法的企业资信评估

高洪波, 马素萍

(南通广播电视大学科研处, 江苏 南通 226006)

**摘要:** 针对基于传统统计学的企业资信评估方法的不足, 提出基于距离判别法的企业资信评估方法, 并用实例通过 Matlab 软件及其相应工具对其进行了计算判别。结果表明, 基于距离判别法的企业资信评估方案, 可快捷、准确、有效地评价企业是否处于破产状态, 为企业资信评估提供可靠的依据。

**关键词:** 马氏距离判别法; 企业资信评估; Matlab

中图分类号: O212

文献标识码: A

文章编号: 1674-7720(2013)03-0076-03

## Research of enterprise credit evaluation based on Mahalanobis distance discrimination method

Gao Hongbo, Ma Suping

(Management Office of Scientific Research, Nantong Radio and TV University, Nantong 226006, China)

**Abstract:** Based on defect of traditional statistics methods of enterprise credit evaluation, this paper put forward a distance discrimination analysis method for the evaluation of enterprise credit, and through the example of Matlab software and its corresponding tool for calculating the discrimination. The results show that: Mahalanobis distance discrimination method for the evaluation of enterprise credit scheme, can be fast, accurate, effective evaluation of whether the enterprise is in a state of insolvency, provides a reliable basis for the evaluation of enterprise credit.

**Key words:** Mahalanobis distance discrimination method; the evaluation of enterprise credit; Matlab

随着我国经济的高速发展, 作为市场经济的金融市场发展速度不断加快, 逐渐与世界经济接轨, 我国经济形态也更趋向于信用经济。资信评估作为市场经济中的监督力量, 在很大程度上可降低信息不对称性, 能够为评价企业的资信水平提供重要参考依据。科学准确的资信评估可以辅助决策, 降低投资风险。因此, 如何提高企业资信评估的准确性和科学性极其重要<sup>[1]</sup>。企业资信评估是以独立经营的企业或经济主体为对象, 根据企业及经济主体的生产、经营、管理前景及当前的企业经济效益状况, 给出企业的资信评级, 本质上是属于综合评价中的分类问题。经济社会活动中判断一个企业是否守信用, 可牵涉到多个数据指标, 如资产负债率、现金流量、流动资产、销售利润率、存贷比、利息偿还率等, 如何从这些数据中判定企业的信用、财务状况, 从而准确地标记出企业的资信等级是一个较为复杂的问题<sup>[2]</sup>。资信评估通常采用基于统计学的分析方法, 包括线性回归分析法、多元判别分析法等。

然而传统的统计学评估方法有较大的局限性, 存在着诸如权重确定缺乏理论依据、带有明显主观臆断且运算量大等缺点, 已经无法满足实际应用的需要。近年来随着科学技术的飞速发展, 尤其是计算机技术的突飞猛进, 基于计算机处理的距离判别法、贝叶斯判别、FISHER 判别等方法在综合评价中的分类问题有了较大的突破和广泛应用。本文将基于 Matlab 实现的距离判别法应用于企业资信评估, 并给出了实例来验证该方法在企业资信评估中财务状况方面评价的可靠性。

### 1 距离判别法原理

判别分析方法是一种有效的多元数据分析方法, 它能从各训练样本中提取各总体的信息, 科学地判断得到的样品属于什么类型, 现已在很多领域广泛应用。马氏距离是由印度统计学家马哈拉诺比斯 (P.C. Mahalanobis) 提出的, 表示数据的协方差距离, 是一种有效的计算两个未知样本集的相似度的方法。

# 技术与方法

## Technique and Method

### 1.1 马氏距离

设  $G$  为  $n$  维总体,其分布的均值向量和协方差矩阵分别为:

$$U=(u_1 \ u_2 \ \dots \ u_n)^T, V=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

设  $x=(x_1, x_2, \dots, x_n)^T, y=(y_1, y_2, \dots, y_n)^T$  为取自总体  $G$  的两个样品,假定  $V>0$  ( $V$  为正定矩阵),定义  $x, y$  间的平方马氏距离为:

$$d^2(x, y) = (x - y)^T V^{-1} (x - y)$$

定义  $x$  到总体  $G$  的平方马氏距离为:

$$d^2(x, G) = (x - U)^T V^{-1} (x - U)$$

### 1.2 马氏距离判别法

可分为两个总体与多个总体判别,实际上多个总体判别可归结为两个总体判别,即多个总体判别可视为两个总体判别的推广。以下介绍两个总体距离判别原理。

设有两个  $n$  维总体  $G_1$  和  $G_2$ , 分布的均值向量分别是  $U_1, U_2$ , 协方差矩阵分别是  $V_1>0, V_2>0$ 。从两个总体中分别抽取容量为  $n_1, n_2$  两个样本, 记为  $x_{11}, x_{12}, \dots, x_{1n_1}$  和  $x_{21}, x_{22}, \dots, x_{2n_2}$ 。现有一未知类别的样品, 记为  $x$ , 则可用以下的判别规则进行判断:

- (1) 若  $d^2(x, G_1) < d^2(x, G_2)$ , 则  $x \in G_1$ ;
- (2) 若  $d^2(x, G_1) > d^2(x, G_2)$ , 则  $x \in G_2$ ;
- (3) 若  $d^2(x, G_1) = d^2(x, G_2)$ , 则待判。

### 1.3 基于 Matlab 的距离判别法的实现<sup>[9-5]</sup>

Matlab 统计工具箱中提供了 `classify` 函数, 用来对未知类别的样品进行判别, 可以进行距离判别和先验分布为正态分布的贝叶斯判别, 其调用格式如下:

`class=classify(sample, training, group, type)`

输入参数 `sample` 是待判别的样本数据矩阵, `training` 是用于构造判别函数的训练样本数据矩阵。它们的每一行对应一个观测值, 每一列对应一个变量, `sample` 和 `training` 具有相同的列数, 该函数将 `sample` 中的每一个观测值归入 `training` 中观测值所在的某个分组。参数 `group` 是与 `training` 相应的分组变量, 参数 `group` 是与 `training` 具有相同的行数, `group` 中的每一个元素指定了参数 `group` 是与 `training` 中相应观测值所在的组。 `group` 可以是一个分类变量、数值向量、字符串数值或者是字符串元胞数组。输出参数 `class` 是一个行向量, 用来指定 `sample` 中各个观测值所在的组, `class` 和 `group` 具有相同的数据类型。参数 `type` 的取值决定了 `classify` 函数支持的判别类型, 其中有五种可选参数, 即 `linear`、`diaglinear`、`quadratic`、`diagquadratic` 及 `mahalanobis`。当 `type` 参数取前四种值时, 该函数可用来作贝叶斯判别; 当取值为 `mahalanobis` 时, 该函数用作距离判别, 并且所计算的先验概率用来计算误判概率。

## 2 企业资信中财务状况评估模型构建<sup>[6]</sup>

### 2.1 企业资信评估指标选取

企业资信评估通过分析独立经营企业资产实力、偿债能力和信用风险程度等, 从而确定该企业的信用等级, 使管理者掌握企业经营状况, 帮助金融机构决策者对企业进行评价和选择。资信评估包括资产评估和信用评估两个方面, 本文选择 4 项评价指标对企业的财务状况进行评价:

X1=现金流量/总债务

X2=净收入/总资产

X3=流动资产/流动债务

X4=流动资产/净销售额

表 1 是相关企业年度财务数据, 其中收集了 21 个破产的企业在破产前两年的年度财务数据, 同时对 25 个财务状况良好的企业也收集了同一时期的财务数据进行距离判别, 找出未判别的四家企业的财务状况的情况, 判别是否处于破产状态。

表 1 相关企业年度财务数据表 example.xls

编号	组别	X1	X2	X3	X4	编号	组别	X1	X2	X3	X4
1	1	-0.45	-0.41	1.09	0.45	26	2	0.32	0.07	4.24	0.63
2	1	-0.56	-0.31	1.51	0.16	27	2	0.31	0.05	4.45	0.69
3	1	0.06	0.02	1.01	0.40	28	2	0.12	0.05	2.52	0.69
4	1	-0.07	-0.09	1.45	0.26	29	2	-0.02	0.02	2.05	0.35
5	1	-0.10	-0.09	1.56	0.67	30	2	0.22	0.08	2.35	0.40
6	1	-0.14	-0.07	0.71	0.28	31	2	0.17	0.07	1.80	0.52
7	1	0.04	0.01	1.50	0.71	32	2	0.15	0.05	2.17	0.55
8	1	-0.07	-0.06	1.37	0.40	33	2	-0.10	-0.01	2.50	0.58
9	1	0.07	-0.01	1.37	0.34	34	2	0.14	-0.03	0.46	0.26
10	1	-0.14	-0.14	1.42	0.43	35	2	0.14	0.07	2.61	0.52
11	1	-0.23	-0.30	0.33	0.18	36	2	0.15	0.06	2.23	0.56
12	1	0.07	0.02	1.31	0.25	37	2	0.16	0.05	2.31	0.20
13	1	0.01	0.00	2.15	0.70	38	2	0.29	0.06	1.84	0.38
14	1	-0.28	-0.23	1.19	0.66	39	2	0.54	0.11	2.33	0.48
15	1	0.15	0.05	1.88	0.27	40	2	-0.33	-0.09	3.01	0.47
16	1	0.37	0.11	1.99	0.38	41	2	0.48	0.09	1.24	0.18
17	1	-0.08	-0.08	1.51	0.42	42	2	0.56	0.11	4.29	0.44
18	1	0.05	0.03	1.68	0.95	43	2	0.20	0.08	1.99	0.30
19	1	0.01	0.00	1.26	0.60	44	2	0.47	0.14	2.92	0.45
20	1	0.12	0.11	1.14	0.17	45	2	0.17	0.04	2.45	0.14
21	1	-0.28	-0.27	1.27	0.51	46	2	0.58	0.04	5.06	0.13
22	2	0.51	0.10	2.49	0.54	47	未知	-0.16	-0.10	1.45	0.51
23	2	0.08	0.02	2.01	0.53	48	未知	0.41	0.12	2.01	0.39
24	2	0.38	0.11	3.27	0.35	49	未知	0.13	-0.09	1.26	0.34
25	2	0.19	0.05	2.25	0.33	50	未知	0.37	0.08	3.65	0.43

### 2.2 基于 Matlab 的距离判别过程

(1) 读取数据

读取文件 `example.xls` 的第一个工作表中 `C2:F51` 范围的数据, 即全部样本数据, 包括了未判企业数据。

```
>> sample=xlsread('example.xls','C2:F51');
```

《微型机与应用》2013 年第 32 卷第 3 期

## 技术与方法 Technique and Method

读取文件 example.xls 的第一个工作表中 C2:F47 范围的数据,即已知组别的样本数据。

```
>>training=xlsread('example.xls','','C2:F47');
```

读取文件 example.xls 的第一个工作表中 B2:B47 范围的数据,即样本的分组信息数据。

```
>>group=xlsread('example.xls','','B2:B47');
```

列出企业编号

```
>>obs=[1:50];
```

(2)进行马氏距离判别,返回判别结果向量  $C$  和误判概率  $err$

```
>>[C,err]=classify(sample,training,group,mahalanobis);
```

```
>>[obs,C]%查看马氏判别结果
```

得出判别结果如表 2 所示。

表 2 马氏距离判断结果

企业 编号	判断 结果								
1	1	11	1	21	1	31	2	41	2
2	1	12	1	22	2	32	2	42	2
3	1	13	1	23	2	33	2	43	2
4	1	14	1	24	2	34	1	44	2
5	1	15	2	25	2	35	2	45	2
6	1	16	2	26	2	36	2	46	2
7	1	17	1	27	2	37	2	47	1
8	1	18	1	28	2	38	2	48	2
9	1	19	1	29	2	39	2	49	1
10	1	20	1	30	2	40	2	50	2

从表中可以看出,共有 3 个观测值发生了误判,即第 15、16 和 34 号,其中第 15 和第 16 号由第 1 组(财务良好)误判为第 2 组(破产企业),而第 34 号观测值原本属于第 2 组误判为第 1 组,用  $P(j|i)$  来表示原本属于第  $I$  组的样品被误判为第  $J$  组的概率,则误判概率的估计值分别是:

$$P(2|1)=2/21=0.095 \quad P(1|2)=1/25=0.04$$

假设两组的先验概率均为 0.5,则 `classify` 函数的误判概率是:

$$err=0.5 \cdot P(2|1)+0.5 \cdot P(1|2)=0.0676$$

可见该马氏判别的结果是可以令人接受的且基本是合理的。表 2 中的第 47~50 号企业是未判企业的观测值,即未知组别的样品。由以上结果可知,第 47 号和第 49 号企业的观测值判归第 1 组,从而判定它们是破产企业,第 48 号和第 50 号企业的观测值被判归为第 2 组,它们为非破产的财务状况良好的企业。

本文提出将基于距离判别算法用于企业资信评估,该方法具有以下优点:(1)与资信评估常用的统计学方法不同,该方法无需事先建立数学模型,只需将从各训练样本中提取各总体的信息,科学地判断得到的样品属于什么类型,评价过程方便、快捷。(2)不需要人为确定权重,从而避免由于评价过程中的主观因素所导致的结果失真。(3)使用基于距离判别算法,即便出现误差,也可通过假设的先验概率计算误判概率,得出较为客观的结论,其评价结果比传统的统计学方法更为客观、有效。

### 参考文献

- [1] 刘重才,周洲,梅强.LVQ 神经网络在企业资信评估中的应用[J].集团经济研究,2007(10):50-51.
- [2] 李翀,夏鹏.后验概率支持向量机在企业信用评级中的应用[J].计算机仿真,2008,25(5):256-258.
- [3] 苏金明,张莲花,刘波.Matlab 工具箱应用[M].北京:电子工业出版社,2004.
- [4] 董维国.深入浅出 Matlab7.X 混合编程[M].北京:机械工业出版社,2006.
- [5] 张润楚.多元统计分析[M].北京:科学出版社,2006.
- [6] 郑志勇.金融数量分析[M].北京:航空航天大学出版社,2009.

(收稿日期:2012-11-17)

### 作者简介:

高洪波,男,1965 年生,副教授,硕士研究生,主要研究方向:数学建模与数据挖掘。

马素萍,女,1965 年生,高级讲师,主要研究方向:数据挖掘。