

文章编号: 1001-0920(2004)02-0179-04

## 电子商务环境下基于模糊 ISODATA 聚类法的企业风险分类

黄敏, 李凤娥, 孙艺红, 王兴伟  
(东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

**摘要:** 根据电子商务环境下企业风险的特点, 在建立企业风险模糊描述机制的基础上, 基于模糊 ISODATA 聚类法, 对企业风险管理的关键问题——企业风险分类进行研究。实例表明该方法是有用的。

**关键词:** 电子商务; 企业风险; 风险分类; 模糊聚类

中图分类号: TP273 文献标识码: A

## Fuzzy ISODATA cluster based risk sorting for enterprise under EC environments

HUANG Min, LI Feng-e, SUN Yi-hong, WANG Xing-wei

(School of Information Science and Engineering, Northeastern University, Shenyang 110004, China  
Correspondent: HUANG Min, E-mail: mhuang@mail.neu.edu.cn)

**Abstract:** The characteristics of risk for EC enterprise are analyzed. Enterprise risk sorting, which is one of the key problems of risk management under EC environments, is studied by using fuzzy ISODATA cluster based on fuzzy describing of risks. Case study suggested the effectiveness of the method.

**Key words:** electronic commerce(EC); enterprise risk; risk sorting; fuzzy cluster

### 1 引言

随着计算机和通讯技术的发展以及全球化市场经济的形成, 基于 Internet 的电子商务已成为企业运行的发展趋势和必然选择<sup>[1,2]</sup>。电子商务环境下的企业管理成为管理领域的一个热点问题<sup>[3~6]</sup>。在电子商务环境下, 企业运行中存在着更多不确定、不易把握的因素, 如不同地区、地域乃至国家的企业联盟的动态性, 新技术发展所导致的企业发展的超前性等。企业所面临的风险更为突出, 企业风险管理作为一类特殊的管理, 在电子商务环境下的地位显得尤为突出。

风险分类是风险管理的一项基本工作, 它是根

据不同标准对已知风险进行分类, 其目的是深入、全面地认识各项风险, 并有针对性地进行管理。本文针对电子商务下企业风险模糊的特点, 以企业的项目所包含的子过程作为分类标准, 建立了风险的描述模糊机制, 并基于模糊 ISODATA 聚类法, 对电子商务环境下的风险进行模糊分类, 为风险评价与控制奠定了基础。

### 2 风险的模糊描述机制

令分类的风险对象的集合为  $A = \{A_1, A_2, \dots, A_n\}$ , 因素集合为  $B = \{B_1, B_2, \dots, B_m\}$ 。其中:  $n$  为风险的数目,  $m$  为因素的数目。则每个风险对象  $A_i$  的描述向量为

收稿日期: 2002-10-18; 修回日期: 2002-12-27

基金项目: 国家自然科学基金资助项目(70101006, 60003006); 国家高技术研究发展计划资助项目(2001AA 121064); 辽宁省自然科学基金资助项目(20032019, 20032018); 辽宁省博士启动基金资助项目(001015); 教育部留学回国人员科研启动基金资助项目。

作者简介: 黄敏(1968—), 女, 福建长乐人, 教授, 博士, 从事企业风险管理、优化模型算法等研究; 李凤娥(1978—), 女, 甘肃酒泉人, 硕士, 从事虚拟企业风险管理的研究。

$$x_i = x_{i1}, x_{i2}, \dots, x_{im}. \quad (1)$$

### 3 模糊 ISODATA 聚类分析

模糊 ISODATA 聚类是基于模糊表述的模糊分类方法<sup>[7,8]</sup>. 普通分类的聚类判据是使目标函数

$$J(R, V) = \sum_{i=1}^n \sum_{h=1}^c r_{hi} |x_i - V_h|^2 \quad (2)$$

达到极小值 其中

$$r_{hi} = \begin{cases} 1, & \text{对象 } i \text{ 属于分类 } h; \\ 0, & \text{对象 } i \text{ 不属于分类 } h; \end{cases}$$

$c$  为分类数,  $R_{c \times n}$  为分类矩阵,  $V_{c \times m}$  为一组聚类中心向量,  $V_h$  表示第  $h$  类的聚类中心,  $|x_i - V_h|$  表示样本  $x_i$  与聚类中心向量  $V_h$  的距离

聚类准则为求得适当的分类矩阵  $R$  和一组聚类中心向量  $V$ , 使得目标函数最小

模糊分类的聚类判据是使目标函数

$$J(R, V) = \sum_{i=1}^n \sum_{h=1}^c (r_{hi})^q |x_i - V_h|^2 \quad (3)$$

达到极小值 其中:  $R_{c \times n}$  为模糊分类矩阵,  $r_{hi} \in [0, 1]$ ,  $q$  为大于 1 的参数 当  $q = 1$  时, 式(3) 则蜕变为普通分类

聚类准则为求得适当的模糊分类矩阵  $R$  和聚类中心向量  $V$ , 使得目标函数最小

#### 3.1 模糊 ISODATA 聚类步骤

1) 数据规格化: 选用标准化法

$$x_{ik} = \frac{x_{ik} - \bar{x}_i}{\sigma_i}, \quad (4)$$

式中

$$\sigma_i = \sqrt{\frac{1}{m-1} \sum_{k=1}^m (x_{ik} - \bar{x}_i)^2},$$

$$\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$$

2) 确定分类数  $c$  ( $2 \leq c \leq n$ ): 取一初始模糊分类矩阵  $R^{(0)} = M_{fc}$ , 其中

$$M_{fc} = \left\{ R_{c \times n} \mid r_{hi} \in [0, 1], \forall h, i; \sum_{h=1}^c r_{hi} = 1, \forall i; r_{hi} > 0, \forall h \right\}$$

令  $l = 0$ , 转 3).

3) 对于  $R^{(l)}$ , 计算聚类中心向量

$$V^{(l)} = (V_1^{(l)}, V_2^{(l)}, \dots, V_c^{(l)})^T, \quad (5)$$

式中

$$V_h^{(l)} = \frac{\sum_{i=1}^n (r_{hi}^{(l)})^q x_i}{\sum_{i=1}^n (r_{hi}^{(l)})^q},$$

$q$  为大于 1 的正数, 并转 4).

4) 修正模糊分类矩阵  $R^{(l)}$ , 即

$$r_{hi}^{(l+1)} = \frac{1}{\sum_{t=1}^c \left( \frac{|x_i - V_h^{(l)}|}{|x_i - V_t^{(l)}|} \right)^{\frac{2}{q-1}}}, \quad (6)$$

并转 5).

5) 比较  $R^{(l)}$  和  $R^{(l+1)}$ , 若取定的误差精度  $\epsilon > 0$ , 则有

$$\max \left\{ \left| r_{hi}^{(l+1)} - r_{hi}^{(l)} \right| \right\} \leq \epsilon \quad (7)$$

$R^{(l+1)}$  和  $V^{(l)}$  即为所求, 停止迭代; 否则,  $l = l + 1$ , 转 3).

#### 3.2 模糊 ISODATA 聚类效果评价

鉴别聚类效果的评价可采用分类系数指标

$$F_c(R) = \frac{1}{n} \sum_{i=1}^n \sum_{h=1}^c r_{hi}^2 \quad (8)$$

当  $R = M_c$  时,  $F_c(R) = 1$ . 因此  $F_c(R)$  越接近 1, 最终分类的模糊性越小, 聚类效果也越好. 其中

$$M_c = \left\{ R_{c \times n} \mid r_{hi} \in \{0, 1\}, \forall h, i; \sum_{h=1}^c r_{hi} = 1, \forall i; r_{hi} > 0, \forall h \right\}$$

### 4 实例仿真

#### 4.1 实例描述

需要分类的风险对象的集合为

$A =$

{需求风险, 竞争风险, 溢出风险, 市场波动风险, 经济滑坡风险, 金融风险, 环境风险, 政策风险, 质量风险, 成本风险, 时间风险, 技术风险, 沟通风险, 技术衔接风险, 技术外泄风险, 数据质量风险, 信息系统安全风险, 组织和管理风险, 信用风险, 流动性风险, 激励风险, 战略柔性丧失风险, 投资套牢风险, 投资到位风险}.

分类数和考虑的因素分别为

$$c_1 = 11, m_1 = 40, c_2 = 17, m_2 = 45$$

#### 4.2 仿真结果分析

在算法实现过程中, 本文尝试了两种方法:

方法 1: 在给定初始模糊分类矩阵  $R^{(0)}$  的基础上进行聚类;

方法 2: 在给定初始聚类中心向量  $V^{(0)}$  的基础上进行聚类

影响聚类效果的相关因素包括分类数  $c$ , 初始模糊分类矩阵  $R^{(0)}$ , 误差精度  $\epsilon$  和参数  $q$

##### 4.2.1 不同方法对聚类效果的影响

取误差精度  $\epsilon = 0.00001$ , 参数  $q = 1.45$ , 两种方法的聚类效果对比列于表 1.

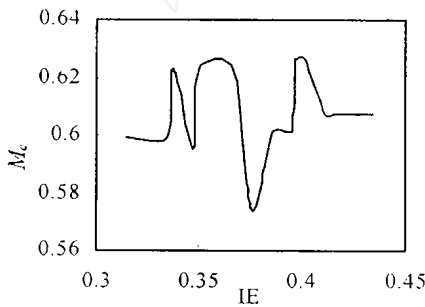
表 1 两种方法的聚类效果对比

分类数	聚类效果评价	
	方法 1	方法 2
$c_1 = 11$	0.601126	0.243985
$c_2 = 17$	0.780527	0.101113

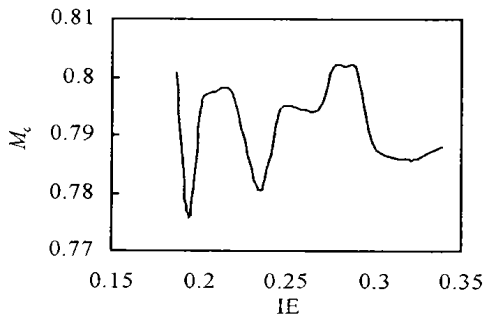
由表 1 可以看出, 方法 1 的聚类效果优于方法 2, 这是因为在选取初始模糊分类矩阵时, 考虑了相关约束条件

4.2.2 不同初始模糊分类矩阵对聚类效果的影响

聚类效果评价价值 ( $M_c$ ) 随初始模糊分类矩阵评价价值 (IE) 变化的曲线如图 1 所示 (误差精度  $\epsilon = 0.00001$ , 参数  $q = 1.45$ ).



(a)  $c = 11$



(b)  $c = 17$

图 1 聚类效果随初始分类矩阵变化曲线

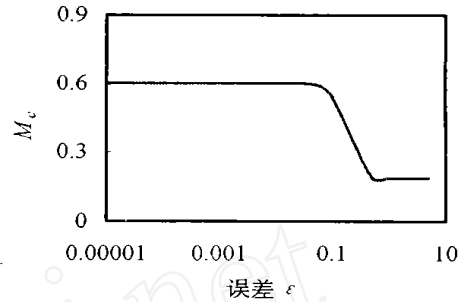
由图 1 可见, 初始分类矩阵对聚类效果的影响很不稳定 因此在确定初始分类矩阵时, 应主要考虑项目的实际风险及其特征

4.2.3 不同误差精度对聚类效果的影响

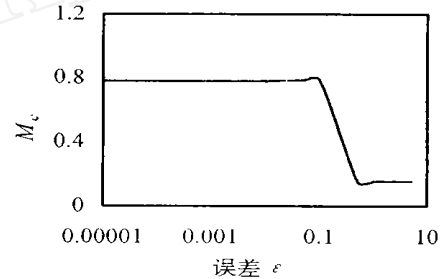
聚类效果随误差精度  $\epsilon$  变化的曲线如图 2 所示 (参数  $q = 1.45$ ). 由图可见, 当误差精度  $\epsilon < 0.01$  时, 聚类可取得满意的效果 这符合精度越高聚类效果越好, 精度越低聚类效果越差的实际情况

4.2.4 不同参数 q 对聚类效果的影响

聚类效果随参数  $q$  变化的曲线如图 3 所示 由图可见, 聚类效果随参数  $q$  的增大而变差 参数  $q$  越接近 1, 分类的模糊性越小, 当  $q = 1$  时, 分类变成硬

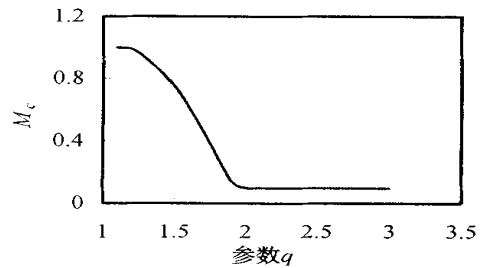


(a)  $c = 11$

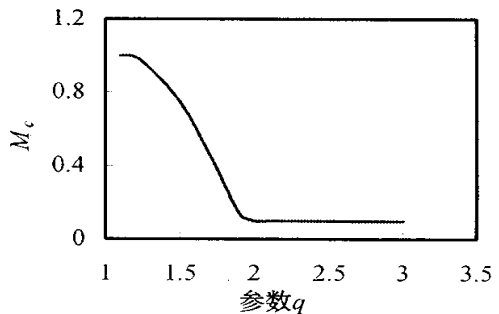


(b)  $c = 17$

图 2 聚类效果随误差精度变化曲线



(a)  $c = 11$



(b)  $c = 17$

图 3 聚类效果随参数的变化曲线

分类; 参数  $q$  越大, 分类的模糊性越大, 它的意义也更不明确 根据实际情况, 本文选择  $q = 1.45$

由上述分析可以看出, 分类数对聚类效果的影响并不明显, 但随着分类数的增大, 程序的运行时间会增大. 当  $c = 11$  时, 运行时间为 3 s; 当  $c = 17$  时, 运行时间为 4 s

4.2.5 分类结果

根据 4.2.1~ 4.2.4 的分析, 采用分类数  $c = 11$ , 误差精度  $\epsilon = 0.00001$ , 参数  $q = 1.45$ . 当初始分

类矩阵评价值为 0.386458 时, 最终分类结果列于表 2

由于对样本集初始矩阵进行了标准化处理, 在

表 2 分类结果

$c$	1	2	3	4	5	6	7	8	9	10	11
1	0.00005	0.99955	0.00002	0.00002	0.00009	0.00004	0.00004	0.00004	0.00007	0.00004	0.00004
2	0.17134	0.21221	0.04138	0.03548	0.11356	0.06250	0.06167	0.05687	0.16065	0.05456	0.02978
3	0.92739	0.00700	0.00297	0.00523	0.01089	0.00827	0.01276	0.00425	0.01099	0.00737	0.00288
4	0.17986	0.07280	0.03363	0.04749	0.12518	0.14992	0.05799	0.07223	0.17191	0.06493	0.02406
5	0.05137	0.06989	0.01861	0.02341	0.37800	0.05206	0.03707	0.04958	0.12381	0.04930	0.14690
6	0.00000	0.00001	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00000	0.99995
7	0.00006	0.00002	0.00010	0.99940	0.00007	0.00008	0.00010	0.00005	0.00006	0.00004	0.00002
8	0.00000	0.00000	0.99999	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
9	0.05437	0.03192	0.03682	0.04486	0.13992	0.09657	0.09726	0.20909	0.16097	0.09186	0.03636
10	0.06524	0.03491	0.03010	0.04088	0.19246	0.06786	0.07249	0.10337	0.29883	0.06398	0.02988
11	0.00044	0.00020	0.00014	0.00034	0.00149	0.00130	0.00192	0.99207	0.00135	0.00054	0.00021
12	0.05138	0.01118	0.00669	0.02443	0.06815	0.08666	0.51411	0.13975	0.06400	0.02496	0.00869
13	0.01829	0.00520	0.00469	0.00828	0.02835	0.83461	0.03298	0.02023	0.01643	0.02739	0.00355
14	0.00484	0.00094	0.00080	0.00331	0.00385	0.00901	0.96009	0.01030	0.00339	0.00273	0.00074
15	0.00756	0.00112	0.00094	0.00460	0.00577	0.00744	0.95239	0.00782	0.00631	0.00517	0.00088
16	0.03325	0.01849	0.01734	0.04138	0.11116	0.22466	0.14304	0.09315	0.08211	0.22107	0.01435
17	0.04207	0.01412	0.02615	0.21130	0.10464	0.08048	0.15135	0.12206	0.10669	0.13044	0.01070
18	0.00049	0.00015	0.00013	0.00021	0.00154	0.00101	0.00047	0.00040	0.00084	0.99461	0.00015
19	0.03260	0.01277	0.01440	0.04253	0.16578	0.31176	0.07046	0.06851	0.19229	0.07535	0.01355
20	0.00882	0.00265	0.00360	0.00901	0.02095	0.88637	0.01677	0.01733	0.01607	0.01490	0.00353
21	0.00429	0.00187	0.00093	0.00160	0.93912	0.00605	0.00312	0.00542	0.02548	0.01083	0.00129
22	0.87663	0.00266	0.00277	0.00546	0.02311	0.01662	0.02030	0.01197	0.02759	0.01057	0.00232
23	0.00442	0.00127	0.00135	0.00132	0.02344	0.00356	0.00281	0.00461	0.95140	0.00400	0.00182
24	0.03006	0.00878	0.00927	0.01551	0.36882	0.18518	0.04435	0.08095	0.18081	0.05793	0.01834

迭代运算的整个过程中, 使得分类阵始终满足模糊划分空间  $M_{fc}$  的 3 个条件. 从表 2 可以看出:

$$r_{hi} = 1, \quad r_{hi} > 0, \quad r_{hi} \in [0, 1]$$

因此模型是正确的, 分类结果符合客观实际

5 结 语

本文针对电子商务环境下企业风险管理的关键问题——企业风险分类进行研究, 首先根据电子商务下企业风险的特点, 建立了企业风险模糊描述机制; 然后基于模糊 ISODATA 聚类法, 对企业风险进行分类; 最后通过实例分析了参数对聚类效果的影响, 表明了该方法的有效性

参考文献(References):

[1] Schneider G P, Perry J T. *Electronic Commerce* [M]. Cambridge: Course Technology, 2000

[2] Kalakota R. *Electronic Commerce: A Manager's Guide*

[M]. Reading: Addison-Wesley, 1997.

[3] Anonymous. Extended ERP meets e-business [J]. *Manufacturing Systems*, 2000, 18(7): 124-154

[4] Gunn J. Extended enterprise integration [J]. *British Telecom Technology J*, 2000, 18(2): 93-99

[5] Tinham B. Switch on to global integrated business and manufacturing [J]. *Manufacturing Computer Solutions*, 2000, 6(9): 46-47.

[6] Grefen P, Aberer K, Hoffner Y, et al. Crossflow: Cross-organizational workflow management in dynamic virtual enterprises [J]. *Computer Systems Science and Engineering*, 2000, 15(5): 277-290

[7] Bezdek J C. *Pattern Recognition with Fuzzy Objective Function Algorithms* [M]. New York: Plenum, 1981. 65-93

[8] 黄洪钟. 模糊设计 [M]. 北京: 机械工业出版社, 1999. 68-77.