

文章编号:1001-9081(2005)10-2447-03

## 集值信息系统知识发现和灰理论在 DRP 中的应用

钟 响,张洪伟,翁怀荣,罗 英  
(四川大学 计算机学院,四川 成都 610064)  
(turezx@tom.com)

**摘 要:**探讨了如何利用灰理论中的灰关联度改善集值信息系统知识发现中包含度的计算,并将其应用在一大型乳业集团的 DRP 系统中来预测客户资信评估的可信度,采用 ASP+XML 和 MS SQL Server 2000 数据库实现,为该集团评估客户的资信提供了科学的方法。

**关键词:**知识发现;包含度;灰理论;灰关联度;分销资源计划

**中图分类号:** TP18 **文献标识码:** A

## Knowledge discovery of sets information system and grey theory and theirs application in DRP

ZHONG Xiang, ZHANG Hong-wei, WENG Huai-rong, LUO Ying  
(School of Computer Science, Sichuan University, Chengdu Sichuan 610064, China)

**Abstract:** How to use the degree of grey incidence of grey theory to improve the calculation of the inclusion degree in sets information system was discussed. This algorithm was implemented in the DRP (Distribution Resource Planning) of a large dairy farming corporation to forecast the reliability of customers' credit. It is designed with the use of asp & xml and MS SQL Server 2000 database and provides a scientific method on evaluating credit ranks of the corporation's customers.

**Key words:** knowledge discovery; inclusion degree; grey theory; degree of grey incidence; Distribution Resource Planning (DRP)

DRP 是企业利用分销管理的先进理念,并结合信息技术,对自己的分销渠道及相关的资源进行管理、调配、监控和优化。客户资信评价的判别方法和模型层出不穷,但迄今为止还没有公认的、有效的和统一的方法。本文基于知识发现提出了一种新的评定方法,并结合灰关联度给出了结果的可信度,即根据已知客户的资信情况评定新客户的资信情况并给出可信度。结合 DRP 系统在该集团的应用,本文所述方法的实用性和科学性得到验证。

### 1 集值信息系统与知识发现

#### 1.1 包含度

包含度理论<sup>[1]</sup>是一种描述不确定性关系的有效度量方法,在人工智能、专家系统和模糊集理论等领域有着重要的应用。包含度的概念如下:

设  $(X, \leq)$  是偏序集,若对于任意  $x, y \in X$  有数  $D(y/x)$  对应,且满足:

- 1)  $0 \leq D(y/x) \leq 1$ ,
- 2) 当  $x \leq y \Rightarrow D(y/x) = 1$
- 3) 当  $x \leq y \leq z \Rightarrow D(x/z) \leq D(x/y)$

则称  $D$  为  $X$  上的包含度<sup>[2]</sup>。

#### 1.2 集值信息系统和目标集值信息系统

称  $(U, A, F)$  为集值信息系统,若  $U = \{x_1, x_2, \dots, x_n\}$  为对象集,  $A = \{a_1, a_2, \dots, a_m\}$  为属性集,  $F = \{f_l: l \leq m\}$  为对象属性值映射,也称信息函数,其中:

$$f_l: U \rightarrow P_0(V_l) \quad (l \leq m)$$

$V_l$  是属性  $a_l$  的值域,  $P_0(V_l)$  表示  $V_l$  的非空子集全体。

记:

$$V = \bigcap_{l=1}^m P(V_l) = \{E = \{E_1, E_2, \dots, E_m\} : E_l \in P(V_l) \quad (l \leq m)\}.$$

$$\text{对于任意} \begin{cases} E = \{E_1, E_2, \dots, E_m\} \in V \\ F = \{F_1, F_2, \dots, F_m\} \in V \end{cases}$$

" $E \supseteq F$ " 表示  $E_l \supseteq F_l, (\forall l \leq m)$ 。于是  $(V, \supseteq)$  是一个偏序集。在  $(V, \supseteq)$  上定义运算:

$$E \vee F = (E_1 \cup F_1, E_2 \cup F_2, \dots, E_m \cup F_m)$$

$$E \wedge F = (E_1 \cap F_1, E_2 \cap F_2, \dots, E_m \cap F_m)$$

则  $(V, \supseteq)$  是一个格,其中最大元和最小元分别为:

$$1 = (V_1, V_2, \dots, V_m), 0 = (\varphi, \varphi, \dots, \varphi)$$

对于集值信息系统  $(U, A, F)$ , 记:

$$D = \{E(x) = (f_1(x), f_2(x), \dots, f_m(x)) : x \in U\}$$

则显然有  $D \in V$ 。

对于集值信息系统  $(U, A, F)$ , 始终假定  $1 = (V_1, V_2, \dots, V_m) \notin D$ 。设  $M \subseteq D$ , 用  $H(M)$  表示  $M$  中不含 1 的由  $M$  生成的极大元全体, 则  $H(M)$  满足以下性质:

1)  $H(M)$  是反链, 即对于任何  $E, F \in H(M)$ , 且  $E \neq F$ ,  $E$  与  $F$  是不可比较的, 也即  $E \supseteq F$  与  $F \supseteq E$  都不成立。

2)  $M \supseteq H(M)$ , 即对于任意  $E \in M$ , 存在  $F \supseteq H(M)$ , 使  $E \supseteq F$ 。

收稿日期:2005-04-07; 修订日期:2005-07-10 基金项目:四川金财科技集团资助项目

作者简介:钟响(1981-),男,湖南人,硕士研究生,主要研究方向:智能信息系统、数据库与计算机网络; 张洪伟(1955-),男,四川人,教授,博士,主要研究方向:智能信息系统、数据库与计算机网络; 翁怀荣(1980-),男,四川人,硕士研究生,主要研究方向:智能信息系统、数据库与计算机网络; 罗英(1981-),女,山西人,硕士研究生,主要研究方向:智能信息系统、数据库与计算机网络。

3) 对于任意  $E \in H(M), E \neq 1$ 。

4) 若  $M_1 \subseteq D, M_2 \subseteq D$  则  $M_1 \supseteq M_2 \Rightarrow H(M_1) \supseteq H(M_2)$ 。

称  $(U, A, F, D, G)$  为目标集值信息系统, 如果  $(U, A, F)$  是集值信息系统, 且  $G = \{g_j: j \leq p\}$ , 其中  $g_j: U \rightarrow V_j$ 。

设  $(U, A, F, D, G)$  为目标集值信息系统, 记:

$$R_A = \{(x, y) \in U \times U: f_i(x) \supseteq f_i(y) (\forall a_i \in A)\}$$

$$R_d = \{(x, y) \in U \times U: g_i(x) = g_i(y) (\forall j \leq p)\}。$$

若  $R_A \subseteq R_d$ , 称目标集值信息系统  $(U, A, F, D, G)$  是协调的。记:

$$U/R_d = \{D_1, D_2, \dots, D_r\},$$

$$M_j = \{E(x) \in D: x \in D_j\} (j \leq r)。$$

对于协调的目标集值信息系统  $(U, A, F, D, G)$  而言,  $\{H(M_j): j \leq r\}$  具有以下性质:

1)  $M_j \supseteq H(M_j) (j \leq r)$ 。

2)  $H(M_j) = H(\{E(y): y \in \bigcup_{x \in D_j} [x]_A\})$ 。

3)  $H(M_j)$  是反链。

4)  $D \supseteq \bigcup_{j=1}^r H(M_j)$

5)  $M_i \cap M_j = \varnothing (i \neq j)$ 。

### 1.3 目标集值信息系统的知识发现

对于一般的目标集值信息系统而言, 以下两种情况下无法给出决策规则:

1) 存在  $E \in V$ , 使  $\{E\} \not\supseteq H(M_j) (\forall j \leq r)$ 。

2) 存在  $j_1, j_2 \leq r (j_1 \neq j_2)$  与  $E \in V$ , 使  $\{E\} \supseteq H(M_{j_1})$  且  $\{E\} \not\supseteq H(M_{j_2})$ 。为此引进包含度方法。

$$\text{设 } \begin{cases} E = \{E_1, E_2, \dots, E_m\} (E_l \subseteq V_l, l \leq m) \\ F = \{F_1, F_2, \dots, F_m\} (F_l \subseteq V_l, l \leq m) \end{cases}$$

记:

$$D(F/E) = \sum_{l=1}^m \frac{|F_l|}{\sum_{l=1}^m |F_l|} X_{F_l}(E_l) \quad (1)$$

其中  $|F_l|$  表示  $F_l$  中的元素个数, 而:

$$x_{F_l}(E_l) = \begin{cases} 1, E_l \subseteq F_l \\ 0, E_l \not\subseteq F_l \end{cases} \quad (2)$$

则  $D(F/E)$  具有以下性质:

1)  $0 \leq D(F/E) \leq 1$ ;

2) 当  $E \leq F$  即  $E_l \subseteq F_l (l \leq m)$  时,  $D(F/E) = 1$ ;

3) 若  $E \leq F \leq G$ , 则  $D(E/G) \leq D(E/F)$

$D(F/E)$  是包含度, 于是可得到知识挖掘规则。设:

$$H(M_j) = \{F_1, F_2, \dots, F_k\}$$

对于任意的  $E = (\{v_1\}, \{v_2\}, \dots, \{v_m\})$ , ( $v_l \in V_l, l \leq m$ ), 计算:

$$D(M_j) = \max_{i \leq k} D(F_i/E), j \leq r$$

$$\text{若 } D(M_{j_0}) = \max_{i \leq k} (D(M_j))$$

则得到规则:

$$(a_1, a_2, \dots, a_m) = (v_1, v_2, \dots, v_m) \Rightarrow d = j_0$$

其可信度为  $D(M_{j_0})$ 。

## 2 灰色系统理论与灰关联度

传统的集值信息系统知识发现包含度的计算可以更加精确, 即更加精确计算  $X_{F_l}(E_l)$  的值, 故引入灰关联度以改善其计算。

### 2.1 灰色系统理论

基于信息思维, 自然现象往往是灰的。灰色现象里含有已知的、未知的与非确知的种种信息; 含有含糊不清的机理;

存在数据不足的表现。少数据与少信息带来的不确定性, 称为灰色不确定性<sup>[3]</sup>。

灰色系统理论(简称灰理论或灰论, Grey Theory<sup>[4]</sup>), 是研究少数据不确定性的理论。灰理论在分析少数据特征、了解少数局的行为表现、探讨少数据的潜在机制、综合少数据现象基础上, 揭示少数据、少信息背景下事物的演化规律提供支持。

### 2.2 灰关联度

灰关联度为一种描述序列之间联系之紧密程度的数量表征<sup>[5]</sup>, 其具有整体性测量的特性。

#### 2.2.1 灰关联因子集

令  $X$  为序列集,  $x_i$  为序列:

$$X = \{x_i | i \in I = \{1, 2, \dots, w\}\}, x_i = (x_i(1), x_i(2), \dots, x_i(n))$$

若  $x_i$  满足下述条件, 则称  $X$  为灰关联因子集:

1) 序列  $x_i$  的数据  $x_i(k)$  之间具有数值可比性。即指定  $x_i(k)$  与  $x_j(t)$  之间数值相等或接近、或同数量级。

2) 序列  $x_i$  之间具有可接近性, 即序列之间具有非平行性。

3) 序列  $x_i$  同极性, 即序列之间的极性应该统一(为正极性)。

#### 2.2.2 差序列

在灰关联因子集  $X$  中, 指定  $x_0$  (或  $x_i$ ) 为参考序列,  $x_j$  为比较序列, 则参考序列对于各比较序列间的绝对差记为  $\Delta_{ij}(k)$ :

$$\Delta_{ij}(k) = |x_i(k) - x_j(k)|, k = 1, 2, \dots, n, j \in J = \{1, 2, \dots, m\}$$

称  $\Delta_{ij}$  为差序列:

$$\Delta_{ij} = (\Delta_{ij}(1), \Delta_{ij}(2), \dots, \Delta_{ij}(n))$$

#### 2.2.3 边界值

称下述差值为边界值:

$$\text{上边界值 } \Delta(\max) = \max_i \max_j \max_k \Delta_{ij}(k)$$

$$\text{下边界值 } \Delta(\min) = \min_i \min_j \min_k \Delta_{ij}(k)$$

当只有一个参考序列  $x_i = x_0$  时, 则有:

$$\Delta_j(k) = |x_0(k) - x_j(k)|$$

$$\Delta_j = (\Delta_j(1), \Delta_j(2), \dots, \Delta_j(n))$$

$$\Delta(\max) = \max_j \max_k \Delta_j(k)$$

$$\Delta(\min) = \min_j \min_k \Delta_j(k)$$

#### 2.2.4 灰关联系数

令  $r(x_i(k), x_j(k))$  为比较序列  $x_j$  对于参考序列  $x_i$  在第  $k$  点的灰关联系数, 则其算式为:

$$r(x_i(k), x_j(k)) = \frac{\min_i \min_j \min_k \Delta_{ij}(k) + \xi \max_i \max_j \max_k \Delta_{ij}(k)}{\Delta_{ij}(k) + \xi \max_i \max_j \max_k \Delta_{ij}(k)} \quad (3)$$

上式中,  $\xi$  为分辨系数,  $\xi \in [0, 1]$ , 一般取  $\xi = 0.5$ 。

#### 2.2.5 灰关联度

$r(x_i, x_j)$  为  $x_j$  对于  $x_i$  的灰关联度:

$$r(x_i, x_j) = \frac{1}{n} \sum_{k=1}^n r(x_i(k), x_j(k)) \quad (4)$$

## 3 客户资信评估

客户资信评估, 包括资产和信用两个方面的评估, 它是对独立经营企业或经济主体进行客户资产实力、偿债能力和信用风险程度的分析, 确定对受信客户的信用条件, 对社会营造一个良好的经营环境。

#### 4 实验数据和评价结果

首先我们构建客户信用等级评定信息系统的数学模型。其中:

$U = \{x_1, x_2, \dots, x_{10}\}$ ,  $x_1, x_2, \dots, x_{10}$  表示 10 个客户;

$A = \{a_1, a_2, a_3, a_4\}$ ; 其中  $a_1$  表示企业经营者素质属性,  $a_2$  表示企业财务状况属性,  $a_3$  表示社会信用状况属性,  $a_4$  表示企业发展前景属性,  $Va_1 = Va_2 = Va_3 = Va_4 = \{\{1\}, \{2\}, \dots, \{10\}\}$ ;

$D = \{d\}$ ;  $d$  表示评价结果;

$Vd = \{0, 1, 2, 3, 4, 5\}$ ;

根据集团现有客户资信情况经过专家分析建立如表 1 的系统。

表 1 集值目标信息系统

$U$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$x_1$	{8}	{10}	{10}	{9}	5
$x_2$	{3}	{4}	{5}	{4}	2
$x_3$	{6}	{7}	{6}	{6}	3
$x_4$	{2}	{3}	{2}	{2}	1
$x_5$	{1}	{1}	{1}	{1}	0
$x_6$	{8}	{5}	{7}	{7}	3
$x_7$	{6}	{6}	{5}	{6}	3
$x_8$	{3}	{6}	{8}	{6}	2
$x_9$	{9}	{8}	{9}	{9}	4
$x_{10}$	{4}	{6}	{2}	{3}	2

得到已知规则:

$D_0 = \{x_5\}, D_1 = \{x_4\}, D_2 = \{x_2, x_8, x_{10}\}, D_3 = \{x_3, x_6, x_7\}, D_4 = \{x_9\}, D_5 = \{x_1\}$

$M_0 = \{(\{1\}, \{1\}, \{1\}, \{1\})\}, M_1 = \{(\{2\}, \{3\}, \{2\}, \{2\})\},$

$M_2 = \{(\{3\}, \{4\}, \{5\}, \{4\}), (\{3\}, \{6\}, \{8\}, \{6\}), (\{4\}, \{6\}, \{2\}, \{3\})\},$

$M_3 = \{(\{6\}, \{7\}, \{6\}, \{6\}), (\{8\}, \{5\}, \{7\}, \{7\}), (\{6\}, \{6\}, \{5\}, \{6\})\},$

$M_4 = \{(\{9\}, \{8\}, \{9\}, \{9\})\}, M_5 = \{(\{8\}, \{10\}, \{10\}, \{9\})\},$

$H(M_0) = \{(\{1\}, \{1\}, \{1\}, \{1\})\}, H(M_1) = \{(\{2\}, \{3\}, \{2\}, \{2\})\},$

$H(M_2) = \{(\{3, 4\}, \{4, 6\}, \{2, 5, 8\}, \{3, 4, 6\})\},$

$H(M_3) = \{(\{6, 8\}, \{5, 6, 7\}, \{5, 6, 7\}, \{6, 7\})\},$

$H(M_4) = \{(\{9\}, \{8\}, \{9\}, \{9\})\}, H(M_5) = \{(\{8\}, \{10\}, \{10\}, \{9\})\}.$

一新客户资信指标为  $(5, 6, 5, 6)$ 。

$E = \{\{5\}, \{6\}, \{5\}, \{6\}\}$ , 根据 (1)、(2) 式则:

$D(M_0) = 0, D(M_1) = 0, D(M_2) = 0.8,$

$D(M_3) = 0.8, D(M_4) = 0, D(M_5) = 0$

于是有规则:

$(5, 6, 5, 6) = (a_1, a_2, a_3, a_4) \Rightarrow d = 2(0.8)$

$(5, 6, 5, 6) = (a_1, a_2, a_3, a_4) \Rightarrow d = 3(0.8)$

可见两条新规则的可信度相同, 无法最终确定新客户的资信, 所以引入灰关联度以确定其可信度。改善公式 (2) 如下:

$$x_{f_i(E_i)} = \begin{cases} 1, & E_i = F_i \\ \max r(x_i, x_j), & E_i \subset F_i \\ 0, & E_i \not\subset F_i \end{cases}$$

$x_i, x_j$  分别是参考序列和比较序列, 即由已知规则构成参考序列, 待定规则构成比较序列。

本例中,  $(a_1, a_2, a_3, a_4) \Rightarrow d = 2$  的参考序列为

$(3, 4, 5, 4), (3, 6, 8, 6), (4, 6, 2, 3);$

比较序列为  $(5, 6, 5, 6);$

初始化为  $(1, 4/3, 5/3, 4/3), (1, 2, 8/3, 2), (1, 3/2, 1, 3/4);$  和  $(1, 6/5, 1, 6/5)。$

得到差序列:

$\Delta_1 = (0, 2/15, 2/3, 2/15), \Delta_2 = (0, 4/5, 5/3, 4/5),$

$\Delta_3 = (0, 3/10, 1/2, 9/20),$

$\max_i \max_k \Delta_{ij}(k) = 5/3, \min_i \min_k \Delta_{ij}(k) = 0;$

计算  $r(x_i(k), x_j(k))$  分别为:

$(1, 25/27, 5/9, 25/27), (1, 25/49, 1/3, 25/49), (1, 25/34, 5/8, 50/77);$

得到  $r(x_i, x_j)$  为:

$0.851852, 0.588435, 0.752412$

故  $\max r(x_i, x_j) = 0.851852$

最后得到  $D(M_2) = 0.68$

同理,  $(a_1, a_2, a_3, a_4) \Rightarrow d = 3$  的参考序列为:

$(6, 7, 6, 6), (8, 5, 7, 7), (6, 6, 5, 6);$

比较序列为  $(5, 6, 5, 6);$

初始化为  $(1, 7/6, 1, 1), (1, 5/8, 7/8, 7/8), (1, 1, 5/6, 1);$  和  $(1, 6/5, 1, 6/5);$

得到差序列:

$\Delta_1 = (0, 1/30, 0, 1/5), \Delta_2 = (0, 23/40, 1/8, 13/40), \Delta_3 = (0, 1/5, 1/6, 1/5),$

$\max_i \max_k \Delta_{ij}(k) = 23/40, \min_i \min_k \Delta_{ij}(k) = 0;$

计算  $r(x_i(k), x_j(k))$  分别为:

$(1, 69/77, 1, 23/39), (1, 1/3, 23/33, 23/49), (1, 23/39, 69/109, 23/39);$

得到  $r(x_i, x_j)$  为:

$0.871461, 0.624923, 0.703129$

故  $\max r(x_i, x_j) = 0.871461$

最后得到  $D(M_3) = 0.70$

$D(M_3)$  比  $D(M_2)$  可信度更高, 该新客户资信情况为  $d = 3$ 。

#### 5 结语

本文针对采用灰关联度改善集值信息系统包含度的计算进行了介绍, 并采用一大型乳业集团的客户的资信指标数据, 建立了一个集值信息系统模型, 通过本文介绍的知识发现的方法实现了对客户的资信等级的评定和可信度的计算。只要有一定数量的客户资信资料, 就可应用本方法对客户进行评价, 为企业提供客户分析的量化依据。

参考文献:

- [1] GRZYMALA-BUSSE JW. Manageing Uncertainty in Expert Systems [M]. Kluwer Academic Publisher, 1991.
- [2] 张文修, 梁怡, 吴伟志. 信息系统与知识发现 [M]. 北京: 科学出版社, 2003.
- [3] 邓聚龙. 灰预测与灰决策 [M]. 武汉: 华中科技大学出版社, 2002.
- [4] DENG JL. The foundation of grey system theory [M]. The Journal of Grey System, 1997(1): 40.
- [5] 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.