

# 挖掘同对象信息元的传导知识

杨春燕, 蔡文

(广东工业大学 可拓工程研究所, 广东 广州 510090)

**摘要:** 传导知识是由于传导变换的发生而产生的一类规则知识. 由于传导变换包括同对象信息元间和异对象信息元间的传导变换, 因此传导知识也有不同的形式. 这类知识对于人们解决矛盾问题、为决策者提供决策参考很有价值. 以可拓学中的传导变换理论为依据, 首先给出传导特征和传导度等基本概念, 然后探讨变换关于同对象信息元传导知识的表示和获取步骤, 为从数据库中挖掘基于可拓变换的传导知识提供了可操作的方法, 也为研究获取异对象信息元间的传导知识提供参考.

**关键词:** 可拓学; 信息元; 传导度; 传导知识; 可拓数据挖掘

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 1673-4785 (2008) 04-0305-04

## Mining conductive knowledge in information elements possessed by an identical object

YANG Chun-yan, CAI Wen

(Research Institute of Extension Engineering, Guangdong University of Technology, Guangzhou 510090, China)

**Abstract:** Conductive knowledge is a sort of rule knowledge caused by conductive transformation. Conductive knowledge has different forms because conductive transformations vary, including variations between transformations of information elements possessed by a single object and by different objects. This kind of knowledge is valuable for solving contradiction problems and can provide considerable help for decision makers. Based on the conductive transformation theory of Extension, the basic conceptions of conductive characteristics and conductive degree are presented. The procedures for representation and acquisition of conductive knowledge among information-elements possessed by an identical object are discussed. The study provides practical methods for mining conductive knowledge, based on extension transformations from databases, and methodologies for acquisition of conductive knowledge from information elements possessed by different objects.

**Keywords:** extenics; information-element; conductive degree; conductive knowledge; extension data mining

一个可拓变换作用于某一信息元, 会使同对象不同特征的某些信息元产生传导变换<sup>[1]</sup>, 类似的现象经常发生. 那么, 这一传导变换对哪些特征起传导作用, 对哪些特征不起传导作用, 作用有多大, 是量变还是质变, 如果能从数据库中找到这些规律性的知识, 就可为决策者提供决策的依据.

传导变换往往通过很多天 (或多次、多人) 的数据表现出来. 因此, 要了解这些主动变换的传导作

用, 就必须通过参变量 (如时间、对象等) 信息元来发现它们的变化与实现的主动变换有无关系以及关系又有多大.

本文将以前文<sup>[1-2]</sup>中的传导变换理论<sup>[3-4]</sup>为依据, 研究传导特征和传导度等基本概念, 探讨变换关于同对象信息元的传导知识的表示和获取步骤. 这类知识在已有的数据挖掘方法和技术<sup>[5-6]</sup>中还未涉及.

## 1 基本概念

### 1.1 变换关于同对象信息元传导度的概念 给定多维信息元

收稿日期: 2008-04-25.

基金项目: 国家自然科学基金资助项目 (70671031); 广东省普通高校人文社会科学研究重点资助项目 (06ZD63008); 广东省自然科学基金资助项目 (05001832).

通信作者: 杨春燕. E-mail: ywy@gdut.edu.cn

$$I(t) = \begin{bmatrix} O(t) & c_1 & v_1(t) \\ O(t) & c_2 & v_2(t) \\ \dots & \dots & \dots \\ O(t) & c_m & v_m(t) \end{bmatrix} = \begin{bmatrix} I_1(t) \\ I_2(t) \\ \dots \\ I_m(t) \end{bmatrix}$$

若在  $t_0$  时刻有主动变换, 使对某一信息元  $I_{j_0}(t_0) = (O(t_0), c_{j_0}, v_{j_0}(t_0))$ , 有

$$I_{j_0}(t_0) = I_{j_0}(t) = (O(t), c_{j_0}, v_{j_0}(t)), t > t_0,$$

且  $v_{j_0}(t) = a \cdot v_{j_0}(t_0)$ , 其中  $a$  为常数, 即主动变换实施后, 该信息元的量值不再随时间而变化. 则  $I_{j_0}$  的传导变换  $I_{j_0} T_{I_{j_0}}$  使

$$I_j(t_0) = (O(t_0), c_j, v_j(t_0)),$$

$$j = j_0, j = \{1, 2, \dots, m\}$$

变为

$$I_{j_0} T_{I_{j_0}} I_j(t_0) = I_j(t) = (O(t), c_j, v_j(t)).$$

式中:  $j = j_0, j = \{1, 2, \dots, m\}, t > t_0$ .

$$\text{称 } j(t) = \frac{v_j(t) - v_j(t_0)}{|v_{j_0}(t) - v_{j_0}(t_0)|} = \frac{v_j(t) - v_j(t_0)}{|a \cdot v_{j_0}(t_0)|} = \frac{v_j}{|v_{j_0}|}$$

为关于  $I_j$  的传导度.  $j(t) > 0$ , 表示实施后, 对  $I_j$  的传导变换  $I_{j_0} T_{I_{j_0}}$  使  $I_j$  的量值  $v_j$  增加;  $j(t) < 0$ , 表示量值  $v_j$  减少;  $j(t) = 0$ , 表示量值不变.

上述多维信息元可用信息元集表示为

$$\{I_j(t)\} = \{(O(t), c_j, v_j(t)), t > t_0\}, j = 1, 2, \dots, m.$$

若数据库中有多个时刻  $t_1, t_2, \dots, t_q$  的数据, 则

$$I_j(t_p) = (O(t_p), c_j, v_j(t_p)),$$

$$j = \{1, 2, \dots, m\}, p = \{1, 2, \dots, q\}, t_p > t_0.$$

取

$$j_{\min} = \frac{\min_{1 \leq p \leq q} \{v_j(t_p) - v_j(t_0)\}}{|v_{j_0}|},$$

$$j_{\max} = \frac{\max_{1 \leq p \leq q} \{v_j(t_p) - v_j(t_0)\}}{|v_{j_0}|},$$

称  $[j_{\min}, j_{\max}]$  为关于  $I_j$  的传导度区间.  $j_{\min} > 0$ , 表示对  $I_j$  的传导变换使  $I_j$  的量值  $v_j$  增加;  $j_{\max} < 0$  表示对  $I_j$  的传导变换使  $I_j$  的量值  $v_j$  减少.

为叙述方便, 把变换关于信息元集

$$\{I_j(t_p)\} = \{(O(t_p), c_j, v_j(t_p)), p = 1, 2, \dots, q, t_p > t_0\}$$

的传导度视为对特征  $c_j (j = \{1, 2, \dots, m\})$  的传导度.

例如, 对 (产品  $O(t_0)$ , 价格,  $v(t_0)$ ) 实施变换后变为 (产品  $O(t)$ , 价格,  $v(t)$ ), 则变换关于信息元集 { (产品  $O(t)$ , 销售量,  $v(t)$ ) } 的传导度视为对“销售量”这一特征的传导度.

## 1.2 非传导特征和传导特征

对信息元集

$$\{I_j(t)\} = \{(O(t), c_j, v_j(t)), t > t_0\},$$

$$j = \{1, 2, \dots, m\},$$

及  $t_0$  时刻的某一信息元

$$I_{j_0}(t_0) = (O(t_0), c_{j_0}, v_{j_0}(t_0)),$$

若变换使

$$I_{j_0}(t_0) = (O(t), c_{j_0}, v_{j_0}(t)) = I_{j_0}(t),$$

且对任一  $t > t_0$ ,  $v_{j_0}(t) = a \cdot v_{j_0}(t_0)$ , 即主动变换实施后的量值  $v_{j_0}(t)$  不再随  $t$  的变化而变化, 而成为一常数  $a$  且关于  $\{I_j(t)\}$  的传导变换  $I_{j_0} T_{I_{j_0}}$  使

$$I_{j_0} T_{I_{j_0}} I_j(t_0) = I_j(t) = (O(t), c_j, v_j(t)).$$

当  $t > t_0$  时, 传导度

$$j(t) = \frac{v_j(t) - v_j(t_0)}{|a \cdot v_{j_0}(t_0)|} = 0$$

则称  $c_j$  为  $I_{j_0}$  的非传导特征.  $c_j$  的非传导特征的全体构成  $I_{j_0}$  的非传导特征集, 记作

$$C^-(c_j) = \{c_j | v_j(t) - v_j(t_0) = 0,$$

$$j = j_0, j = 1, 2, \dots, m\}.$$

若原特征集  $C = \{c_1, c_2, \dots, c_n\}$ , 称

$$C(c_j) = C - C^-(c_j) - \{c_{j_0}\}$$

为  $I_{j_0}$  的传导特征集.

显然, 对于  $c_j \in C(c_j)$ , 至少存在一个  $t_1 > t_0$ , 使  $v_j(t_1) \neq v_j(t_0)$ .

## 2 变换关于同对象信息元的传导知识

根据上述概念, 可以从数据库或数据仓库中挖掘变换关于同对象信息元的传导知识, 即从数据库中获取如下知识.

### 2.1 关于变换的传导特征集和非传导特征集

$$[I_{j_0}(t_0) = I_{j_0}(t)] \Rightarrow (l_1) C(c_j),$$

$$\text{式中: } l_1 = \left\lceil q \cdot \frac{|I_{j_0}(t_p)|}{q} \right\rceil.$$

$$[I_{j_0}(t_0) = I_{j_0}(t)] \Rightarrow (l_2) C^-(c_j),$$

$$\text{式中: } l_2 = \left\lceil q \cdot \frac{q - |I_{j_0}(t_p)|}{q} \right\rceil, |I_{j_0}(t_p)| \text{ 是传导}$$

信息元集

$$\{I_j(t_p)\} = \{I_j(t_p) | v_j(t_p) - v_j(t_0) = 0,$$

$$t_p > t_0, p = 1, 2, \dots, q\}, j = \{1, 2, \dots, m\}$$

中的信息元个数,  $q$  为  $t_p$  的个数.

### 2.2 传导特征对应量值的变化范围

对信息元

$$I_j(t) = (O(t), c_j, v_j(t)), j = \{1, 2, \dots, m\}$$

在数据库中有多个时刻的数据, 记为

$$I_j(t_p) = (O(t_p), c_j, v_j(t_p)),$$

$j \in \{1, 2, \dots, m\}, p \in \{1, 2, \dots, q\},$   
记  $t_0$  为原始时刻,对某一信息元

$$I_{j_0}(t_0) = (O(t_0), c_{j_0}, v_{j_0}(t_0))$$

实施主动变换：

$$I_{j_0}(t_0) = I_{j_0}(t) = (O(t), c_{j_0}, a),$$

且  $a = v_{j_0}(t_0).$

记  $a_j = \min_{1 \leq p \leq q} \{v_j(t_p)\}, b_j = \max_{1 \leq p \leq q} \{v_j(t_p)\},$ 取

$$v_j = [a_j, b_j], j \in \{1, 2, \dots, m\}.$$

对信息元  $I_j(t) = (O(t), c_j, v_j(t)),$ 若  $c_j$  为传导特征,则有

$$[I_{j_0}(t_0) = I_{j_0}(t)] \Rightarrow (l_1)(v_j(t) = V_j).$$

式中：
$$l_1 = \left[ q, \frac{| \{ I_j(t_p) \} |}{q} \right].$$

2.3 变换关于传导特征的传导度

若  $I_j(t)$  满足条件  $L: v_j(t) = V_j,$  则有

$$(I_{j_0}(t_0) = I_{j_0}(t)) \quad (I_j(t) = L) \Rightarrow$$

式中：
$$l_1 = \left[ \frac{| \{ I_j(t_p) \} |}{q} \right]_{j \in \min, j \in \max} J).$$

3 挖掘变换关于同对象信息元传导知识的步骤

3.1 列出变换 实施前后的原始信息元表

将变换 实施前后的原始数据按照信息元的要求规范整理,列成如表 1 的原始信息元表.

3.2 计算变换实施前后原始信息元量值差表

计算变换 实施前后的原始元的量值之差,列成表 2

表 1 变换 实施前后的原始信息元表

Table 1 Original information elements before and after actualizing a transformation

	$c_1$	...	$c_{j_0}$	...	$c_j$	...	$c_m$
$O(t_0)$	$v_1(t_0)$	...	$v_{j_0}(t_0)$	...	$v_j(t_0)$	...	$v_m(t_0)$
$O(t_1)$	$v_1(t_1)$	...	$a$	...	$v_j(t_1)$	...	$v_m(t_1)$
$O(t_2)$	$v_1(t_2)$	...	$a$	...	$v_j(t_2)$	...	$v_m(t_2)$
...	...	...	...	...	...	...	...
$O(t_q)$	$v_1(t_q)$	...	$a$	...	$v_j(t_q)$	...	$v_m(t_q)$

表 2 变换 实施前后原始信息元量差值表

Table 2 The difference values of original information-elements fore-and-aft of actualizing a transformation

	$c_1$	...	$c_{j_0}$	...	$c_j$	...	$c_m$
$O(t_1)$	$v_1(t_1) - v_1(t_0)$	...	$a - v_{j_0}(t_0)$	...	$v_j(t_1) - v_j(t_0)$	...	$v_m(t_1) - v_m(t_0)$
$O(t_2)$	$v_1(t_2) - v_1(t_0)$	...	$a - v_{j_0}(t_0)$	...	$v_j(t_2) - v_j(t_0)$	...	$v_m(t_2) - v_m(t_0)$
...	...	...	...	...	...	...	...
$O(t_q)$	$v_1(t_q) - v_1(t_0)$	...	$a - v_{j_0}(t_0)$	...	$v_j(t_q) - v_j(t_0)$	...	$v_m(t_q) - v_m(t_0)$

3.3 列出传导特征对应的原始信息元

根据表 2 的数据,再根据传导特征的规定,确定传导特征,并将传导特征对应的原始信息元列成表 3

3.4 计算变换 关于各特征的传导度区间

利用前述传导度计算公式计算各信息元的传导度和传导特征的传导度区间,列成表 4

表 3 传导特征对应的原始信息元表

Table 3 The original information elements corresponding with conductive characteristics

传导特征	各对象对应于传导特征的量值				最小最大量值		量值区间 $v_j$
	$O(t_1)$	$O(t_2)$	...	$O(t_q)$	$\min v_j(t_p)$	$\max v_j(t_p)$	
$c_1$	$v_1(t_1)$	$v_1(t_2)$	...	$v_1(t_q)$	$a_1$	$b_1$	$[a_1, b_1]$
...	...	...	...	...	...	...	...
$c_j$	$v_j(t_1)$	$v_j(t_2)$	...	$v_j(t_q)$	$a_j$	$b_j$	$[a_j, b_j]$
...	...	...	...	...	...	...	...
$c_m$	$v_m(t_1)$	$v_m(t_2)$	...	$v_m(t_q)$	$a_m$	$b_m$	$[a_m, b_m]$

表 4 各信息元的传导度和传导特征的传导度区间

传导特征	各对象对应于传导特征的传导度				最小最大传导度		传导度区间
	$O(t_1)$	$O(t_2)$	...	$O(t_q)$	$j_{\min}$	$j_{\max}$	
$c_1$	$1(t_1)$	$1(t_2)$	...	$1(t_q)$	$1_{\min}$	$1_{\max}$	$[1_{\min}, 1_{\max}]$
...	...	...	...	...	...	...	...
$c_j$	$j(t_1)$	$j(t_2)$	...	$j(t_q)$	$j_{\min}$	$j_{\max}$	$[j_{\min}, j_{\max}]$
...	...	...	...	...	...	...	...
$c_m$	$m(t_1)$	$m(t_2)$	...	$m(t_q)$	$m_{\min}$	$m_{\max}$	$[m_{\min}, m_{\max}]$

3.5 获取有关的传导知识

从 3.3 和 3.4 可以获得如下关于同对象信息元的传导知识:

3.5.1 变换的传导特征集和非传导特征集

$I_{t_0}(t_0) = (O(t), c_0, a) \Rightarrow (l)C^-(c_j)$ ,  
式中  
 $C^-(c_j) = \{c_j \mid v_j(t) - v_j(t_0) = 0, \\ j = j_0, j = 1, 2, \dots, m, t > t_0\},$   
 $I_{t_0}(t_0) = (O(t), c_0, a) \Rightarrow (l)C(c_j)$ . (1)  
式中:  $C(c_j) = C - C^-(c_j) - \{c_0\}$ , 它表示变换的  
传导特征集是  $C(c_j)$ , 非传导特征集是  $C^-(c_j)$ . 式  
(1) 中,  $l = \left[ q, \frac{|I_j(t_0)|}{q} \right]$ ,  $q$  是样本数,  
 $\frac{|I_j(t_0)|}{q}$  是可信度.

根据表 3, 可得传导特征集为

$C(c_j) = \{c_1, \dots, c_j, \dots, c_m\}$   
3.5.2 传导特征对应量值的变化范围  
对  $c_j \in C(c_j)$ , 信息元为  
 $I_j(t) = (O(t), c_j, v_j(t))$ ,  
则有

$I_{t_0}(t_0) = (O(t), c_0, a) \Rightarrow$   
 $(l)\{v_j(t) \mid V_j = [a_j, b_j]\},$   
式中:  $l = \left[ q, \frac{|I_j(t_0)|}{q} \right]$ , 样本数为  $q$ , 可信度为  
 $\frac{|I_j(t_0)|}{q}$ .

3.5.3 变换关于传导特征的传导度和传导度区间  
对  $c_j \in C(c_j)$ , 信息元

$I_j(t) = (O(t), c_j, v_j(t))$   
满足条件  $L: v_j(t) \in [a_j, b_j]$ , 则有  
 $I_{t_0}(t_0) = (O(t), c_0, a) \Rightarrow (I_j(t) \in L) \Rightarrow$   
 $(l)\{j(t) \mid [j_{\min}, j_{\max}]\},$   
式中:  $l = \left[ q, \frac{|I_j(t_0)|}{q} \right]$ ,  $q$  是样本数,  $\frac{|I_j(t_0)|}{q}$   
是可信度.

4 结束语

同对象信息元之间的传导变换是常见的传导变

换, 也是人们解决矛盾问题的常用方法, 同时也是导致新的矛盾问题产生的根源之一. 因此, 研究从数据库中挖掘同对象信息元的传导知识是很有必要的. 本文将为利用历史数据获取传导知识, 为解决矛盾问题提供可操作的方法, 也为研究获取一般信息元的传导知识提供参考.

参考文献:

[1] 杨春燕, 蔡文. 可拓工程 [M]. 北京: 科学出版社, 2007.  
[2] 蔡文, 杨春燕, 何斌. 可拓逻辑初步 [M]. 北京: 科学出版社, 2003.  
[3] 李立希, 杨春燕, 李铎汶. 可拓策略生成系统 [M]. 北京: 科学出版社, 2006.  
[4] YANG Chunyan. Conductive transformation and conductive contradiction problem solving[C] // Proceedings of 2006 International Conference on AI Beijing: BUPT Publishing House, 2006: 840-843.  
[5] 陈文伟. 挖掘变化知识的可拓数据挖掘研究 [J]. 中国工程科学, 2006, 8(11): 70-73.  
CHEN Wenwei. Extension data mining for mining changing knowledge [J]. Engineering Science, 2006, 8(11): 70-73.  
[6] 陈安, 陈宁, 周龙骧, 等. 数据挖掘技术及应用 [M]. 北京: 科学出版社, 2006.

作者简介:



杨春燕, 女, 1964 年生, 研究员, 中国人工智能学会理事, 中国人工智能学会可拓工程专业委员会常务副主任, 主要研究方向为可拓学、人工智能、决策科学. 主持国家自然科学基金项目 2 项, 广东省自然科学基金项目 2 项, 参加国家自然科学基金项目多项, 发表学术论文 50 余篇, 出版专著 7 部.



蔡文, 男, 1942 年生, 研究员, 国家级有突出贡献的专家, 可拓学的创立者, 中国人工智能学会常务理事, 中国人工智能学会可拓工程专业委员会主任, 主要研究方向为可拓学、人工智能、决策科学. 主持国家自然科学基金项目 5 项, 参与国家自然科学基金项目多项, 发表学术论文多篇, 出版专著 7 部.