学术论文

挖掘变化知识的可拓数据挖掘研究

陈文伟
（海军兵种指挥学院，广州 510431）

[摘要] 规范了可拓信息与可拓知识基本概念，即在信息和知识的概念上扩充了变化的信息和变化的知识。明确了可拓数据挖掘概念以及可拓推理新概念。证明了可拓数据挖掘两个定理和可拓推理公式。提出的从挖掘静态知识的数据挖掘扩展到挖掘变化知识的可拓数据挖掘，为数据挖掘开辟了新的研究方向，并通过实例进行了说明。

[关键词] 可拓信息；可拓知识；可拓数据挖掘；可拓推理
[中图分类号] TP18 [文献标识码] A [文章编号] 1009－1742（2006）11－0070－04

1 可拓信息与可拓知识概念

可拓学的理论和方法与信息科学和智能科学交叉融合的研究，具体来说就是通过可拓信息与可拓知识来改变问题的目的或条件，去解决矛盾问题。

1.1 可拓知识是知识概念的扩展

可拓信息是解决矛盾问题的信息。可拓学基元理论中的基元（物元、事元、关系元）是可拓信息的基础信息。可拓学中的可拓变换是变化信息，通过变换才能变矛盾问题为不矛盾问题。

可拓信息 = 基元（基础信息）⊕可拓变换（变化信息）。

目前对信息的定义，基本上属于对信息的静态描述。可拓信息中的基元信息也属于对信息的静态描述，而变换信息属于变化的信息，具有变化特征。

解决矛盾问题必须通过可拓变换，即利用变化的信息才能解决矛盾问题。可见，可拓信息是信息概念的拓展。

知识是对信息进行加工，或者对信息进行浓缩，找出事物中存在的规律，如表达式、蕴含式等。知识概念仍具有静态性。

可拓知识是解决矛盾问题的知识。可拓学的拓展原理的表达式，即拓展式（发散式、相关式、可扩式、蕴含式等）是可拓知识的基础知识。可拓学的传导原理的变换蕴含式是变化知识。可拓学引入关联函数将矛盾问题进行量化处理，称它为量化知识。

可拓知识 = 拓展式（基础知识）⊕
变换蕴含式（变化知识）⊕关联函数，

其中变换蕴含式为 $(T_u u = u') \rightarrow ({}_uT_v v = v')$，简写为 $(T_u \rightarrow {}_uT_v)$。

可拓知识中的拓展式中的蕴含式与人工智能的产生式规则是一致的。拓展式中的发散式、相关式、可扩式等可以看成是产生式的扩展。它们仍具有静态特征。

可拓知识中的变换蕴含式是典型的变化知识。它是解决矛盾问题的更有价值的知识。可见，可拓知识是知识概念的扩展。

1.2 解决矛盾问题的量化知识——关联函数

关联函数公式 $k(x) = \rho(x, x_0, X_0)/D(x, X_0, X)$，其中 $X_0 = \langle a, b\rangle$，$k(x) > 0$ 是正域区间。$X = \langle c, d\rangle$，$k(x) < 0$，是质变区间。

关联函数本身属于知识。当 x 从区间 X_0 变化

[收稿日期] 2006－01－09；修回日期 2006－02－12
[作者简介] 陈文伟（1940－），男，江西南昌市人，海军兵种指挥学院教授，国防科技大学博士生导师

到区间 X 后，即关联函数 $k(x)$ 由正数变为负数，表明矛盾问题得到解决。

2　挖掘变化知识的可拓数据挖掘

数据挖掘是从数据中挖掘出知识。由于数据具有静态性，所挖掘的知识也具有静态性。

2.1　可拓数据挖掘概念

提出可拓数据挖掘，在于挖掘可拓知识，它是数据挖掘的扩展。主要包含如下两类：

1）*挖掘关联函数的区间信息*　解决矛盾问题的量化方法是建立关联函数，通过可拓推理使变量 x 从区间 X_0 变换到 X，区间参数 a，b，c，d 一般是运用实验或统计得到。利用数据挖掘方法，获取区间参数信息，是可拓数据挖掘的一类重要任务。

2）*挖掘变换蕴含式的可拓数据挖掘*　数据挖掘获取知识（条件→结论），对条件进行可拓变换和对结论进行传导变换，获得的变化的知识，即可拓知识

$$T_{条件} \rightarrow T_{结论},$$

把这种挖掘变化的知识称为新型的可拓数据挖掘。

2.2　可拓数据挖掘理论

定理 1　对于两类规则

$$A \rightarrow P \tag{1}$$

$$B \rightarrow N \tag{2}$$

一般情况 $A = \wedge a_i, B = \wedge b_i$。

若存在条件的可拓变换

$$T_{条件}(B) = A \tag{3}$$

并存在结论的可拓变换 $T_{结论}$（它为 $T_{条件}$ 的传导变换），

$$T_{结论}(N) = P \tag{4}$$

则可拓变换规则知识（变化知识）

$$T_B(B) = A \rightarrow T_N(N) = P \tag{5}$$

成立，即　if $T_B(B) = A$ then $T_N(N) = P$　(6)

证明：

1）定理的已知条件表示成命题逻辑公式，并化为子句型

a. $A \rightarrow P \leftrightarrow \neg A \vee P$,

b. $B \rightarrow N \leftrightarrow \neg B \vee N$,

c. $T_B(B) = A \leftrightarrow \neg B \wedge A \leftrightarrow \neg B, A$,

d. $T_N(N) = P \leftrightarrow \neg N \wedge P \leftrightarrow \neg N, P$。

2）对定理的结论取非后化成子句型

$$\neg(T_B(B) = A \rightarrow T_N(N) = P) \leftrightarrow$$
$$\neg[(\neg B \wedge A) \rightarrow (\neg N \wedge P)] \leftrightarrow$$
$$\neg[\neg((\neg B) \wedge A) \vee (\neg N \wedge P)] \leftrightarrow$$
$$\neg[(B \vee \neg A) \vee (\neg N \wedge P)] \leftrightarrow$$
$$\neg(B \vee \neg A) \wedge \neg(\neg N \wedge P) \leftrightarrow$$
$$\neg B \wedge A \wedge (N \vee \neg P) \leftrightarrow$$
$$\neg B, A, N \vee \neg P。$$

3）对全部子句集进行归结

a. 全部子句集为

$\neg A \vee P, \neg B \vee N, \neg B, A, \neg N, P, N \vee \neg P$。

b. 归结过程：子句 $\neg A \vee P$ 与子句 A 归结为 P，它与子句 $N \vee \neg P$ 归结为 N，再和子句 $\neg N$ 归结为空子句，产生矛盾，故证明定理正确。

定理 2　对于两条同类规则

$$A \rightarrow P \tag{7}$$

$$C \wedge B \rightarrow P \tag{8}$$

若存在可拓变换

$$T_B(B) = A \tag{9}$$

则可拓变换规则知识

$$T_B(B) = A \rightarrow P \tag{10}$$

成立，即 if $T_B(B) = A$ then P　(11)

该定理同样可用归结原理证明（略）。

2.3　可拓数据挖掘过程

从可拓数据挖掘定理中，可以概括可拓数据挖掘过程为：

Step 1 对分类问题利用数据挖掘方法获得分类规则，即获得式（1）和式（2）的知识。

Step 2 确定规则的前提中存在的可拓变换以及结论中存在的可拓变换，即找出满足式（3）和式（4）的可拓变换。

Step 3 利用定理 1 和定理 2 获得可拓知识式（5）或式（10）。

3　可拓推理是知识推理的扩展

在智能科学中，知识推理采用了形式逻辑中的假言推理。可拓推理是对拓展式和变换蕴含式的假言推理。

1）拓展推理　对拓展式的假言推理称为拓展推理。以发散式为例，发散式推理表示为

$$(N_1, c_1, v_1) \wedge [(N_1, c_1, v_1) \dashv (N_1, c_1, v_i)] \vdash (N_1, c_1, v_i) \tag{12}$$

2）传导推理　变换蕴含式是可拓变换与传导变换之间的蕴含式，它的假言推理称为传导推理，表示为

$$(T_u u = u') \wedge [(T_u u = u') \rightarrow ({}_u T_v v = v')] \vdash ({}_u T_v v = v') \quad (13)$$

可拓推理是在知识推理的基础上，扩展为对变化知识的推理。

证明：

1）将式（13）中推理（$\vdash$）的左部写成等价的命题逻辑公式

$(\neg u \wedge u') \wedge [(\neg u \wedge u') \rightarrow (\neg v \wedge v')]$。

2）上式化为子句型

$$(\neg u \wedge u') \wedge [(\neg u \wedge u') \rightarrow (\neg v \wedge v')] \leftrightarrow$$
$$(\neg u \wedge u') \wedge [\neg(\neg u \wedge u') \vee (\neg v \wedge v')] \leftrightarrow$$
$$(\neg u \wedge u') \wedge [(u \vee \neg u') \vee (\neg v \wedge v')] \leftrightarrow$$
$$(\neg u \wedge u') \wedge [(u \vee \neg u' \vee \neg v) \wedge (u \vee \neg u' \vee v')] \leftrightarrow (\neg u \wedge u') \wedge$$
$$(u \vee \neg u' \vee \neg v) \wedge (u \vee \neg u' \vee v') \leftrightarrow$$

$\neg u$, u', $(u \vee \neg u' \vee \neg v)$, $(u \vee \neg u' \vee v')$。

3）将推理（$\vdash$）的右部取非后，化为子句型

$\neg(Tv = v') \leftrightarrow \neg(\neg v \wedge v') \leftrightarrow v \vee \neg v'$。

4）归结过程：子句 $v \vee \neg v'$ 与子句 $(u \vee \neg u' \vee \neg v)$ 归结为 $\neg v' \vee u \vee \neg u'$，它与子句 $\neg u$ 归结为 $\neg v' \vee \neg u'$，与 u' 归结为 $\neg v'$，再与子句 $(u \vee \neg u' \vee v')$ 归结为 $u \vee \neg u'$，与 $\neg u$ 归结为 $\neg u'$，再与 u' 归结为空子句。产生矛盾，证明可拓推理式(13)是正确的。

可拓知识只表明存在变化的可能性。可拓推理表明实际变化的发生。在式（5）中，可拓知识 $(T_u \rightarrow T_v)$ 表明对 u 的变换 T_u 会引起对 v 的变换 T_v。而可拓推理式(13)表明现已发生变换 T_u，按式(13)必然出现变换 T_v。

4 可拓数据挖掘与可拓推理实例

在“脑血栓”与“脑出血”两类疾病的数据库中进行数据挖掘和可拓数据挖掘。

4.1 在数据库中通过数据挖掘获取规则知识

从“脑出血”和“脑血栓”两种疾病的大量实例数据库中，通过数据挖掘的遗传算法可以获取两种疾病独立诊断的规则知识。获得的主要7条规则（具体数据挖掘过程略）：

1）（高血压 = 有）∧（瞳孔不等大 = 是）∧（膝腱反射 = 不活跃）→ 脑出血，

2）（瞳孔不等大 = 是）∧（语言障碍 = 是）→脑出血，

3）（高血压 = 有）∧（起病方式 = 快）∧（意识障碍 = 深度）→脑出血，

4）（高血压 = 有）∧（病情发展 = 快）→脑出血，

5）（高血压 = 有）∧（动脉硬化 = 有）∧（起病方式 = 慢）→脑血栓，

6）（动脉硬化 = 有）∧（病情发展 = 慢）→脑血栓，

7）（动脉硬化 = 有）∧（意识障碍 = 无）→脑血栓。

4.2 确定存在的可拓变换

在医疗中病人存在的可拓变换有：

$T_{条件}$（起病方式慢）= 起病方式快，

$T_{条件}$（无意识障碍）= 深度意识障碍，

也存在可拓变换 $T_{结论}$（脑血栓）= 脑出血。

4.3 利用可拓数据挖掘定理 获取可拓知识（变化的知识）

根据定理1得到可拓变换知识（变化知识）为

T(有动脉硬化 ∧ 起病方式慢 ∧ 无意识障碍) = 起病方式快 ∧ 有深度意识障碍

$$\rightarrow T(\text{脑血栓}) = \text{脑出血} \quad (14)$$

还可以得出其他的可拓知识。

4.4 可拓推理

可拓知识中的前提一旦在现实中出现，就可以利用可拓推理判断可拓知识中结论的出现。当发现某病人由起病方式慢变成起病方式快，同时无意识障碍变成有深度意识障碍，即可拓知识式（14）的前提已经出现，利用可拓推理式（13）就可以判断可拓知识式（14）的结论已经出现，即应该诊断该病人已经由“脑血栓”变成了“脑出血”。治疗方式就应改由“脑血栓”的治疗方法变成治疗“脑出血”的方法。

两种疾病的治疗方法是完全相反的，若仍然用“脑血栓”的治疗方法治疗“脑出血”，将会快速加重“脑出血”症状，甚至于导致死亡。这条变化知识对医生来讲是极其重要的。

可见，挖掘变化知识的可拓数据挖掘比挖掘静态知识的数据挖掘更有意义。

参考文献

[1] 陈文伟，黄金才，赵新昱. 数据挖掘技术[M]. 北京：北京工业大学出版社，2002

[2] 蔡 文，杨春燕，何 斌. 可拓逻辑初步[M]. 北京：科学出版社，2003

[3] 陈文伟. 可拓学与智能科学、信息科学[A]. 香山科学会议(第 271 次会议)[C]. 北京：香山科学会议第 271 次学术讨论会筹备组，2005. 7～50

[4] 陈文伟，黄金才. 从数据挖掘到可拓数据挖掘[A]. 中国人工智能进展[C]，北京：北京邮电大学出版社，2005. 844～848

The Research of Mining the Mutative Knowledge With Extension Data Mining

Chen Wenwei

(*Naval Arms Command Academy*, *Guangzhou* 510431, *China*)

[**Abstract**] The thesis standardized the extension information and the extension knowledge, that is, basing on the information and knowledge, the thesis extended the mutative information and knowledge, and nailed down the conceptions of extension data mining and extension reasoning. This thesis also attested two theorems of the extension data mining and the formula of the extension reasoning. The author put forward the extension data mining which extended from mining the static knowledge to mining the mutative knowledge. So it exploited a new aspect of data mining, and illuminated it by examples.

[**Key words**] extension information; extension knowledge; extension data mining; extension reasoning

《中国工程科学》2006 年第 8 卷第 12 期要目预告