基于本体的可拓知识链获取

陈文伟
(海军兵种指挥学院,广东 广州 510431)

摘 要:相对于静态的知识,定义了基于可拓变换的可拓知识,可拓知识是变化的知识. 在可拓知识定理(从知识中获取可拓知识)和可拓推理公式的基础上,证明了基于集合的可拓知识定理和基于本体的可拓知识链定理. 通过实例,在多维层次数据中,获取问题产生原因的可拓知识链.

关键词:可拓变换;可拓知识;集合;本体;可拓知识链
中图分类号:TP18 **文献标识码**:A **文章编号**:1673-4785(2007)06-0068-04

Acquisition of an extensional knowledge chain based on ontology

CHEN Wen-wei
(Naval Arms Command Academy, Guangzhou 510431, China)

Abstract: This paper begins with a definition of extensional knowledge based on extensional transformations. Extensional knowledge is mutative knowledge relative to static knowledge. Based on the theorem of extensional knowledge and the formula for extensional knowledge reasoning, a theorem of extensional knowledge based on sets and a theorem of an extensional knowledge chain based on ontology were proven. Using a case study, the process for generating an extensional knowledge chain in multi-dimensional hiberarchical data was analyzed.

Keywords: extensional transformation; extensional knowledge; set; ontology; extensional knowledge chain

1 可拓变换与可拓知识概念

1.1 可拓变换概念

定义 1 在可拓学中,可拓变换定义为对象 u、$v \in \{M$、A、R、k、$U\}$(即物元、事元、关系元、准则、论域中的任一个对象),将对象 u 改变为对象 v 的变换 T 称为可拓变换. 记作

$$Tu = v. \tag{1}$$

1.2 可拓变换形式逻辑表示

式(1)中将对象 u 变为对象 v,实际上完成了 u 自身变为$\sim u$,并使 v 成为真. 这样,可拓变换可以用形式逻辑表示.

定义 2 可拓变换形式逻辑表示为

$$Tu = v \leftrightarrow\!\!\sim u \wedge v. \tag{2}$$

1.3 可拓知识概念和定义

可拓学中可拓知识包括:发散型知识、相关型知识、蕴含型知识、可扩型知识、共轭型知识和变换的蕴含型知识共 6 种类型. 文中主要研究的可拓知识类型是变换的蕴含型知识.

可拓变换可能由某个条件(原因)产生或者可拓变换会引起某个结果,与可拓变换有关的具有因果关系的规则式,统称为变换的蕴含型知识.

1) 可拓变换 Tu 由某一条件或原因所引起:

$$\text{Condition} \rightarrow T_u u = v. \tag{3}$$

①条件 Condition 可能是某一事实 $F = f$,具体表示为

$$F = f \rightarrow T_u u = v. \tag{4}$$

②条件 Condition 可能是另一个可拓变换 $T_a a = b$,具体表示为

$$T_a a = b \rightarrow T_u u = v. \tag{5}$$

此时,在可拓学中,可拓变换 T_u 称为可拓变换 T_a 的传导变换.

③条件 Condition 可能是一个算子 A 求出变量 X 的值,表示为

收稿日期:2007-05-23.
基金项目:国家自然科学基金资助项目(70671013).

$$A(x) = b \rightarrow T_u u = v. \quad (6)$$

2) 可拓变换 T_a 可能产生一个结果:

$$T_a a = b \rightarrow \text{result}. \quad (7)$$

结果 result 同样可能是一个事实,或者是另一个可拓变换.

定义 3　包含可拓变换的规则知识(因果关系式)称为"变换的蕴含型知识"类型的可拓知识.

在式(3)～(7)中,式(5)是典型的可拓知识的代表形式.

1.4　可拓知识与知识的对比

1) 知识:知识在人工智能中一般表示成规则形式,即

$$P \rightarrow Q.$$

式中:P 与 Q 均为事实(变量的取值),它表示事实 P 是事实 Q 的原因,事实 Q 是事实 P 的结果. 知识只体现了 P 与 Q 的静态关系.

2) 可拓知识:在可拓知识中,规则式中包括了可拓变换,而可拓变换将一对象变换成另一个对象,体现了变化的持点.

公式(5)表示可拓变换 T_a 把 a 变换为 b,引起了另一个可拓变换 T_u 把 u 变成 V,这种可拓知识完全体现了变化的情况,故可拓知识是变化知识,相对而言,人工智能中不涉及变换的知识,是静态知识.

也可以说,可拓知识是知识的推广,是一种更有价值的知识[1-5].

2　可拓知识与可拓推理的基础理论

知识与可拓知识之间是有关系的,作者证明了如下 2 个可拓知识定理.

2.1　可拓知识定理(从知识中获取可拓知识)

定理 1　对于 2 类规则:

$$A \rightarrow P, \quad (8)$$

$$B \rightarrow N, \quad (9)$$

若存在条件的可拓变换

$$T_B B = A, \quad (10)$$

并存在结论的可拓变换

$$T_N N = P, \quad (11)$$

则成立可拓知识

$$T_B B = A \rightarrow T_N N = P. \quad (12)$$

该定理的证明见文献[1].

定理 2　对于 2 条同类规则:

$$A \rightarrow P, \quad (13)$$

$$C \wedge B \rightarrow P, \quad (14)$$

若存在可拓变换

$$T_B B = A, \quad (15)$$

则成立可拓知识

$$T_B B = A \rightarrow P. \quad (16)$$

以上 2 条可拓知识定理,提出了从数据挖掘出的知识中,若存在条件的可拓变换及结论的可拓变换,就能够获得相应的的可拓知识(变化的知识). 变化的知识比静态的知识更有价值.

2.2　可拓知识推理

按照形式逻辑的知识推理定义,即假言推理为

$$P \wedge (P \rightarrow Q) \vdash Q. \quad (17)$$

定义 4　可拓知识的推理定义为

$$(T_u u = u') \wedge [(T_u u = u') \rightarrow (_u T_v v = v')] \vdash (_u T_v v = v'). \quad (18)$$

该公式的正确性证明见文献[1].

3　基于集合的可拓知识

在集合论中有集合蕴含关系,定义如下.

定义 5　若集合 P 和 Q 存在关系 $P \subseteq Q$,对于任意的对象 x,若 $x \in P$,则 $x \in Q$,表示蕴含关系为

$$P \rightarrow Q. \quad (19)$$

由此定义,可以得到定理 3.

定理 3　(基于集合的可拓知识)对于可拓变换 $T_a a = b$ 和可拓变换 $T_e e = f$,若存在集合关系 $a \subseteq e$, $b \subseteq f$,则存在可拓知识:

$$T_a a = b \rightarrow T_e e = f. \quad (20)$$

简写为 $T_a \rightarrow T_e$,并称可拓变换 T_a 与 T_e 是同类可拓变换,即 2 个可拓变换前的对象 $\{a, e\}$ 与 2 个可拓变换后的对象 $\{b, f\}$ 均在各同类集合中.

证明

1) 由于 $a \subseteq e$,由定义 5 可知,存在蕴含关系

$$a \rightarrow e. \quad (21)$$

2) 由于 $b \subseteq f$,同样存在蕴含关系

$$b \rightarrow f. \quad (22)$$

根据定理 1 可知,对于式(21)和(22),存在可拓变换 $T_a a = b$ 和 $T_e e = f$,则存在可拓知识

$$T_a a = b \rightarrow T_e e = f.$$

4　基于本体的可拓知识链

本体(ontology)是目前研究最多的知识表示形式,是共享概念的规范化说明,本体在概念分类层次的基础上,加入了关系、公理、规则来表示概念之间的关系.

定义 6(本体)　本体由概念、关系、函数、公理和实例等 5 类基本元素构成,表示为

$$O: = [C, R, F, A, I]. \quad (23)$$

式中:C 为概念,R 为关系,F 为函数,A 为公理,I 为

实例. 关系 R 有 4 种: subclass-of (或 kind-of, 子类)、part-of (部分)、instance-of (实例) 和 attribute-of (属性).

本体概念树的层次关系主要是 subclass-of 关系,即树的下层概念是上层概念的子集,如图 1 所示.

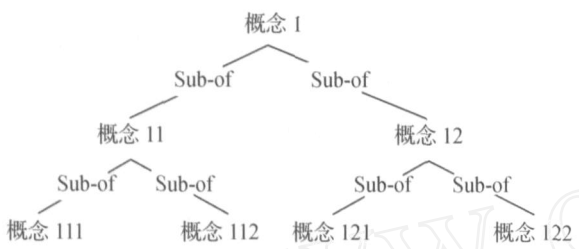

图 1 本体概念树

Fig. 1 Conception tree of ontology

概念 11 的是概念 1 的子集,而概念 111 的是概念 11 的子集,依此类推.

根据本体概念树的特点和定理 3,可以得到定理 4 和定理 5.

定理 4 本体概念层次关系中,下层概念的可拓变换 T_d 与上层概念的同类可拓变换 T_u,存在蕴含关系:

$$T_d \rightarrow T_u. \tag{24}$$

该关系表明 T_d 是主动可拓变换, T_u 是传导可拓变换.

证明 本体概念层次关系中,下层概念集合 S_d 与上层概念集合 S_u 存在蕴含关系: $S_d \subseteq S_u$.

根据定理 3 可知,下层概念集合 S_d 中的可拓变换 T_d 与上层概念集合 S_u 中的同类可拓变换 T_u 存在可拓变换的蕴含关系,即可拓知识:

$$T_d \rightarrow T_u.$$

定理 5 (基于本体的可拓知识链) 在本体概念树中,叶节点中的可拓变换 T_0 与各级上层节点中的同类可拓变换 T_i 之间形成了可拓知识链,即

$$T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \cdots\cdots \rightarrow T_{\text{root}}. \tag{25}$$

证明 由定理 3 可知,本体概念树的上下两层的同类可拓变换都存在蕴含关系 (可拓知识). 由本体概念树叶节点开始,逐层向上到本体概念树的根节点,将同类可拓变换连接起来,就形成式 (25) 的可拓知识链.

可拓知识链表明,根节点的可拓变换 T_{root} 是由叶节点的可拓变换 T_0 引起的传导可拓变换.

5 多维数据中原因分析的可拓知识链获取实例

在我国航空公司数据仓库中,对发现的问题进行原因分析,从中获取可拓知识链. 数据仓库中的多维数据中含层次粒度的大量数据,对发现的问题进行原因分析主要是通过进行多维数据的钻取操作. 在每一次钻取中进行一次可拓变换,获得出现问题原因的深层数据. 数据仓库中的多维层次粒度和数据集合是符合本体概念树的层次关系.

我国航空公司的数据仓库的多维分析中发现了“北京到西南地区总周转量相对去年出现负增长”的问题,该问题的本体概念树如图 2 所示.

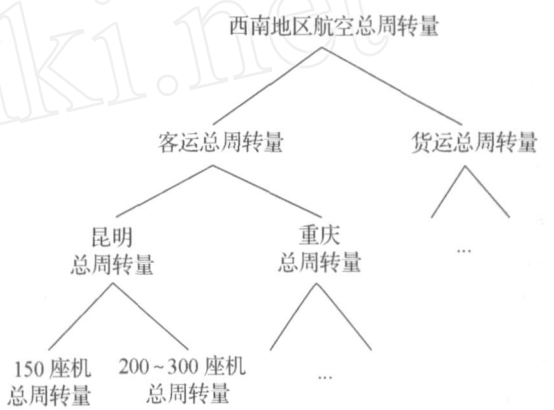

图 2 西南地区航空总周转量的本体概念树

Fig. 2 Conception tree of ontology for overall revolving quantum of southwest airways

该问题在本体树的根节点上的可拓变换表示为

$T_{\text{西南总量}}$ (今年总周转量 - 去年总周转量) = - 19.9 (负增长).

通过下钻到本体树下层,空运总周转量节点上的可拓变换为

$T_{\text{西南客运}}$ (今年客运总周转量 - 去年客运总周转量) = - 19.4 (负增长).

再下钻到昆明客运总周转量节点上的可拓变换为

$T_{\text{昆明客运}}$ (今年总周转量 - 去年总周转量) = - 16.5 (负增长).

再下钻到昆明 150 座机与 200～300 座机的总周转量 2 个结点上的可拓变换分别为

$T_{150\text{座机}}$ (今年总周转量 - 去年总周转量) = - 6.83 (负增长),

$T_{200\sim300\text{座机}}$ (今年总周转量 - 去年总周转量) = - 6.9 (负增长).

根据定理 5,可得到可拓知识链为

$$T_{150\text{座机}} \wedge T_{200\sim300\text{座机}} \rightarrow T_{\text{昆明客运}} \rightarrow T_{\text{西南客运}} \rightarrow T_{\text{西南总量}}.$$

该可拓知识链说明出现西南地区总周转量相对去年出现较大负增长,原因主要是昆明地区 150 座机和 200～300 座机相对去年出现较大负增长造成

的.而该可拓知识链的获得是从问题结论的可拓变换,$T_{西南总量}$出现负增长,通过多维数据钻取,逆向找它的前提可拓变换,再向下钻取,一直到最底层(叶节点)中的可拓变换,$T_{150座机}$及 $T_{200\sim300座机}$出现大的负增长,该叶节点的可拓变换才是本体根节点问题的根本原因.

在向下钻取过程,有时也能发现新问题,如在搜索货运总周转量时,发现东南地区出现了一个大负增长,这是除西南地区出现负增长外新发现的问题,可以在寻找西南地区航空总周转量的根本原因之后,再去寻找东南地区出现货运总周转量出现负增长的原因.

除了寻找负增长以外,还可以寻找正增长的原因.即从正、负 2 个方面寻找问题产生的原因,这样可以得到更大的决策支持.

寻找问题原因让计算机自动完成,必须建立多维层次数据的本体概念树,并在树中进行深度优先搜索,来发现问题并找到所有原因.

6　结束语

数据挖掘是从数据中挖掘知识,可拓数据挖掘是挖掘可拓知识.知识是静态的,而可拓知识是变化的知识.可拓知识定理帮助人们从知识及相关的可拓变换中获取可拓知识.基于本体的可拓知识链定理帮助人们在数据仓库中多维层次数据中获取可拓知识链.目前,数据仓库的问题原因分析基本上是在人的指导下,对多维层次数据进行钻取操作,找到问题发生的原因.若在多维层次数据中建立本体概念树,就可以让计算机沿着本体概念树进行深度优先搜索,既可以发现问题,又能自动找到各问题的所有原因,这项工作是很有意义的.

参考文献:

[1]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. The research of the new technique of mining the mutative knowledge with extension data mining[J]. Engineering Science of China, 2006,8(11):70 - 73.

[2]陈文伟.数据仓库与数据挖掘教程[M].北京:清华大学出版社,2006.

[3]陈文伟,杨春燕,黄金才.可拓知识与可拓知识推理[J].哈尔滨工业大学学报,2006,38(7):1094 - 1096.
CHEN Wenwei, YANG Chunyan, HUANG Jincai. Research on extension knowledge and extension knowledge reasoning[J]. Journal of Harbin Institute of Technology, 2006,38(7):1094 - 1096.

[4]陈文伟,黄金才.从数据挖掘到可拓数据挖掘[J].智能技术,2006,1(2):50 - 52.
CHEN Wenwei, HUANG Jincai. From datamining to extension datamining[J]. Intelligent Technology, 2006,1(2):50 - 52.

[5]杨春燕,蔡　文.可拓工程[M].北京:科学出版社,2007.

作者简介:

陈文伟,男,1940 年生,教授,国防科技大学博士生导师,中国人工智能学会机器学习专业委员会副主任,中国人工智能学会可拓工程专业委员会副主任,主要研究方向为决策支持系统、机器学习、可拓工程、数据仓库与数据挖掘.获国家科技进步奖二等奖 1 项,军队科技进步奖二、三等奖多项.

E-mail:chenww9 @21cn.com.

2008 年 IEEE 计算机科学与软件工程国际会议

2008 International Conference on Computer Science and Software Engineering

2008 International Conference on Computer Science and Software Engineering (CSSE 2008) will be held on December 12～14, 2008 in Wuhan, China. The conference proceedings will be published by IEEE Computer Society, all papers accepted will be included in IEEE Xplore and cited by EI.

Topics of interests include, but not limited to

Computer Science: 1) artificial intelligence; 2) computer graphics; 3) computer application; 4) distributed and parallel computing; 5) database technology; 6) embedded programming; 7) grid computing; 8) information security; 9) scientific computing; 10) human computer interface.

Software Engineering: 1) data mining; 2) project management; 3) requirement analysis; 4) software standards and design; 5) programming methodology; 6) software testing and quality control; 7) web-based services; 8) software maintenance; 9) simulation and modeling.

http://www.highsci.org/csse.