

DOI:10.3969/j.issn.1673-4785.201208037
网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201208037.html>

本体知识拓展分析树在可拓策略生成系统的应用

温树勇, 李卫华
(广东工业大学 计算机学院, 广东 广州 510006)

摘要:针对现有可拓策略生成系统分析矛盾问题难度较大这一问题,利用本体知识和决策树技术共同建立本体知识拓展分析树。首先介绍本体知识拓展分析树模型的建立过程,研究本体知识拓展分析树在可拓策略生成中的应用。然后通过所设计的算法分析本体知识拓展分析树的特征,找出矛盾的核心问题。最后以提高客户价值为例进行实验,实验结果表明,对核心问题的相关特征进行可拓变换,比传统的可拓策略生成系统能更有效生成解决矛盾问题的策略。

关键词:可拓变换;可拓策略生成;本体知识拓展分析树;本体知识;决策树

中图分类号:TP311 **文献标志码:**A **文章编号:**1673-4785(2014)01-0115-06

中文引用格式:温树勇,李卫华.本体知识拓展分析树在可拓策略生成系统的应用[J]. 智能系统学报, 2014, 9(1): 115-120.
英文引用格式:WEN Shuyong, LI Weihua. Application of ontology knowledge expansion analysis tree in the extension strategy generation system[J]. CAAI Transactions on Intelligent Systems, 2014, 9(1): 115-120.

Application of ontology knowledge expansion analysis tree in the extension strategy generation system

WEN Shuyong, LI Weihua
(Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In regard to the phenomenon that it is difficult for the existing extension strategy generating system to analyze a contradiction problem, this paper discusses the ontology knowledge and the decision tree technology being utilized to jointly establish an ontology knowledge expansion analysis tree. The authors present the establishment process for the model of ontology knowledge expansion analysis tree, research the application of the ontology knowledge expanded analysis tree in extension strategy generation, analyze the features of the ontology knowledge expansion analysis tree using the designed algorithm, and find the core problem of contradiction. Finally, an experiment is done with improving a customer's value as example. The results show that, in comparison with the traditional generation system of expandable strategy, carrying out expandable transform for the related characteristics of the core problem may more effectively generate strategies to solve contradiction problems.

Keywords: extension transform; extension strategy generation; ontology knowledge expansion analysis tree; ontology knowledge; decision tree

在可拓学^[1]的实践中,为了协助人们产生解决矛盾问题的策略,创建了一种可拓策略生成系统(extension strategy generating system, ESGS)^[2]。但传统的 ESGS 分析矛盾问题的核心问题时比较困难,产生策略的过程效率不高,因此必须结合其他方

法改进分析过程。

决策树^[3]在分类和预测上有着广泛的应用,包括在客户价值分析中的应用^[4]。本体是共享概念的显式形式化说明^[5],因此可以用可拓学的基元甚至复合元^[6]建立该领域客观事物关系的可拓模型,供计算机存储、检索和处理。

本文将利用本体知识和决策树技术,构造出本体知识拓展分析树,探讨本体知识拓展分析树在可

拓策略生成系统的应用问题,改进可拓策略生成系统分析核心问题的过程。以提高企业的客户价值为研究点,通过决策树技术对客户样本训练集分类,找出分类的实质,即该矛盾的核心问题,以方便后面对该问题应用可拓方法进行可拓变换,达到较快解决矛盾问题的目的。

1 传统可拓策略生成系统分析核心问题的不足

目前,ESGS 通常是先建立可拓模型,即建立问题 P 的核问题 P_0 的可拓模型为 $P_0 = g_0 * l_0$, 其中, g_0 为问题的目标, l_0 为问题的条件。再确定问题的关联函数 K , 然后对问题 P_0 进行相容性分析, 若不相容(目标与条件有矛盾), 则对问题的目标或条件进行可拓分析, 得到问题的发散树或相关树, 通过对问题发散树或相关树的“叶”基元进行可拓变换, 生成候选策略集合。最后对策略集合中的策略进行优度评价, 推荐优度较高的策略供决策者选择。这种对问题发散树或相关树进行可拓变换的方法, 采用的是从“叶子”到“根”, 即从下到上的遍历方法, 存在着以下一些不足之处。

1) 这种采用试探性解决矛盾问题的方法, 需要依次对“叶”基元, 即最小子问题进行可拓变换。当采取某些变换后综合关联函数仍小于零时, 则放弃这些变换(不成为策略), 相当于试错法中发现了错的一步。这种方法不仅增加了程序运行的时间, 而且生成策略针对性不强。

2) 当变换一个“叶”基元满足不了目的时, 需要再对其他“叶”基元进行可拓变换, 而这种可拓变换又可能产生传导变换, 因此, 这种试探性方法容易产生“组合爆炸”的计算复杂性问题, 导致计算机无法解决。

可见, 在 ESGS 中, 有效找到核心问题并对其进行可拓变换至关重要, 使可拓策略生成系统更加智能化。本文将介绍本体知识拓展分析树在可拓策略生成系统中的应用, 首先对训练集生成本体知识拓展分析树, 通过分析该分析树, 有效找出核心问题并对其进行可拓变换生成策略, 并且通过分析树预测变换的效果, 产生准确的策略, 减少策略生成中人为原因产生的误差。

2 本体知识拓展分析树

2.1 本体知识拓展分析树的生成

根据文献[1-2] 阐述: 问题由目标和条件构成, 可用一般模型 $P = G * L$, 其中 G 为目标, 是问题 P 在

一定条件下希望达到的结果, 而 L 为影响目标实现的主客观因素, 即条件。

由于决策树是数据挖掘的有效分类工具, 通过决策树分析原始数据, 可以得到类似“如果 L , 则 G ”的规则知识集。因此, 可以利用本体知识和决策树技术建立本体知识拓展分析树, 通过本体知识拓展分析树分析矛盾问题中的条件 L , 找到影响目标 G 实现的关键因素 l_0 , 建立核问题 $P_0 = g_0 * l_0$ 。

本体知识拓展分析树的生成步骤如下。

1) 根据实际问题, 建立问题领域本体知识模型, 得到相关领域知识。可拓策略生成系统中的本体是用基元/复合元表示。

2) 以本体知识模型为基础, 利用决策树技术, 建立本体知识拓展分析树, 这里的拓展分析主要指相关分析, 但在某些情况下也能用发散、蕴含、可扩分析。根据文献[7]介绍的决策树方法原理, 求出针对所有属性 $c_j (j = 1, 2, \dots, s)$ 的信息增益 $G(c_j)$, 选择信息增益 $G(c_i)$ 最大的属性 c_i 作为决策树的根节点, 并用相同的办法建立子树, 直到递归划分完样本集为止, 从而逐步生成了本体知识拓展分析树。

2.2 本体知识拓展分析树在可拓策略生成中的应用

根据上述建立的本体知识拓展分析树, 分析矛盾问题中的条件, 可以得到影响样本分类的关键节点, 即核心问题。

记样本为信息元 Z_0 , 假设得到一条矛盾问题样本从根到该叶子节点路径为 $K_1: (c_o^1, v_z^1), (c_o^2, v_z^2), \dots, (c_o^i, v_z^i)$, 其中 $c_o^1, c_o^2, \dots, c_o^i$ 为信息元 Z_0 中的任意评价特征, $v_z^1, v_z^2, \dots, v_z^i$ 为评价特征相对应的取值, 对比另外一条相容问题的样本路径 $K_2: (c_o^1, v_z^1), (c_o^2, v_z^2), \dots, (c_o^i, v_z^i)$, 就可以确定导致矛盾问题的关键节点为 c_o^i 。

因此可以通过下面算法得到本体知识拓展分析树中每个样本的路径。

```
FineTreeNode ( Tree t, Node c )
Begin
    If ( t = null ) return OK;
    If ( t->children! = null || t=c ) Return t;
    For ( INT i=0; i<c->childs.length; i++ )
        FineTreeNode ( Tree t->childs[ i ], Node c );
End
```

在一般的实际应用中, 可以根据相关实际, 对特征值分为高、中、低 3 个层次, 因此可以设计以下算法, 得到该路径中的关键节点, 即核心问题 l_0 。

```
CoreProblem ( Node c )
Begin
```

```
If(c.parentArrtribute.equals( middle) ||
    c.parentArrtribute.equals( low) )
    return c→parent
Else
While( c→parent! =null)
Begin
    c=c→parent
    CoreProblem( c→parent)
End
For (Int i=0; i<c→childs.length; i++)
CoreProblem( c→childs[ i] )
End
```

这里,如果该叶子所对应的父节点属性为“中”或“低”时,父节点即核心矛盾,如果叶子所在父节点为“高”时,其核心问题应用同样的方法分析其父节点。

此外,还可以通过决策树对可拓策略的结果进行定性的预测,提高策略的准确性。Forecast(String [] data)的算法如下所示。

```
Forecast( String [ ] data, Node node)
Begin
    If(c→children! =null) Return t;
    Node [ ] childs= node.childNodesArray;
    For (Int i=0; i<c→childs.length;i++)
    Begin
        If ( c→childs[ i]! = null)
        If( c→childs[ i].parentArrtribute.
            equals( printData[ index] ) )
            Forecast( data, c→childs[ i] )
    End
End
```

3 提高客户价值实验

3.1 客户价值案例

文献[8]提出:根据帕累托原则,一个公司 80% 的盈利是由 20% 的客户创造的。从收益管理的角度,并非所有的客户对企业都是同等重要,这就要求企业按照一定的规则对客户进行分类,能够根据不同的客户群体进行有目的性、针对性的投入,从而提高客户价值,减少因为盲目投入造成不必要的浪费。

在对客户群体进行分类时,不仅要参照该客户群体当前的价值表现,而且还需要对客户群体潜在价值进行预测判断。因此,客户价值应该从客户当前价值和潜在价值 2 个方面进行评价^[9]。以提高客户价值为例,问题模型为 $P=G * L$,其中 G 是“提

高客户价值”, L 是“存在潜在价值客户”,要用本体知识拓展分析树分析条件——“潜在价值客户”。

1) 根据文献[10],建立客户价值的本体知识模型,如图 1 所示。根据当前价值和潜在价值进行客户细分,可以得到低价值、潜价值、次价值和价值客户群体^[10]。从本体知识中得知,客户价值与购买数量、品牌忠诚度等属性有关。

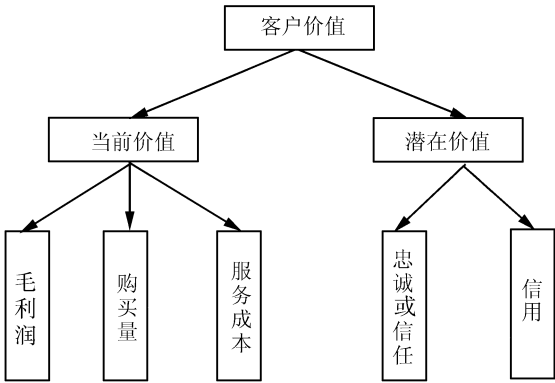


图 1 客户价值的本体知识模型
Fig.1 The ontology model of customer value

2) 根据文献[7]中的方法,利用综合关联函数对原始数据表进行处理,从表 1 中的前 5 个属性可以看出,客户的平均单位产品支付最高为 750 元,最低为 180 元;购买数量最多为 10 件,最少为 1 件;月收入最高为 6 500 元,最低为 2 500 元;品牌忠诚度最高为 5,最低为 1。为简单起见,利用可拓学的简单关联函数和离散关联函数分别建立关于评价特征 $d_{ji}(i=1,2,\cdots,13;j=1,2,3,4)$ 的关联函数 $K_{ji}(x_{ji})$:

$$K_{1i}(x_{1i}) = \frac{x_{1i} - 300}{750 - 180} = \frac{x_{1i} - 300}{570}$$
$$K_{2i}(x_{2i}) = \frac{x_{2i} - 3}{10 - 1} = \frac{x_{2i} - 3}{9}$$
$$K_{3i}(x_{3i}) = \frac{x_{3i} - 4\,000}{6\,500 - 2\,500} = \frac{x_{3i} - 4\,000}{4\,000}$$
$$K_{4i}(x_{4i}) = \begin{cases} 1, & x_{4i} = 5 \\ 0.5, & x_{4i} = 4 \\ 0, & x_{4i} = 3 \\ -0.5, & x_{4i} = 2 \\ -1, & x_{4i} = 1 \end{cases}$$

根据客户价值理论,客户价值分为当前价值 CV 和潜在价值 PV,其关联函数为

$$K_{CV} = (K_{1i}(x_{1i}) + K_{2i}(x_{2i}) + K_{3i}(x_{3i}))/3$$
$$K_{PV} = K_{4i}(x_{4i})$$

因此,可以对客户群体进行分类,得到的关联函数数值和客户分类如表 1 所示。

表 1 关联函数值和客户群体分类

Table 1 Correlation function value and customer groups classification

客户名称	单位产品 支付 d_1 /元	购买数量 d_2 /件	月收入 d_3 /元	品牌忠 诚度 d_4	K_{CV}	K_{PV}	客户分类
客户 O_1	550	7	6 500	4	0.502 680 312	0.5	价值客户
客户 O_2	180	10	4 500	5	0.230 750 487	1	价值客户
客户 O_3	190	1	3 200	1	-0.205 068 226	-1	低价值客户
客户 O_4	450	2	3 300	3	-0.007 651 072	0	潜价值客户
客户 O_5	325	2	6 200	2	0.160 916 179	-0.5	次价值客户
客户 O_6	265	3	3 500	4	-0.062 134 503	0.5	潜价值客户
客户 O_7	220	5	3 000	3	-0.056 042 885	0	潜价值客户
客户 O_8	260	2	2 500	2	-0.185 428 850	-0.5	低价值客户
客户 O_9	375	5	3 900	5	0.109 600 390	1	价值客户
客户 O_{10}	400	3	4 300	2	0.083 479 532	-0.5	次价值客户
客户 O_{11}	280	1	3 500	1	-0.127 436 647	-1	低价值客户
客户 O_{12}	500	3	3 800	1	0.100 292 398	-1	次价值客户
客户 O_{13}	750	8	5 500	4	0.573 343 080	0.5	价值客户

为了计算方便,分别用 M 、 N 、 H 、 L 表示价值客户、次价值客户、潜价值客户和低价值客户 4 类,根据表 1 提供的信息,计算建树所需的信息量:

$$\begin{aligned} \text{Info}(D) = & - \sum_{i=1}^m p_i \text{lb}(p_i) = - \frac{3}{13} \text{lb} \frac{3}{13} - \frac{4}{13} \text{lb} \frac{4}{13} - \\ & \frac{3}{13} \text{lb} \frac{3}{13} - \frac{3}{13} \text{lb} \frac{3}{13} = 1.988 \end{aligned}$$

公司决策层认为:属性单位产品支付 d_1 在 [180, 250) 为低, [250, 450) 为中, [450, 750] 为高;属性购买数量 d_2 在 [1, 4] 为低, [5, 10] 为高;属性月收入 d_3 在 [2 500, 4 000) 为低, [4 000, 6 500] 为高;属性忠诚度 d_4 在 {1, 2} 为低, 3 为中, {4, 5} 为高。因此,属性 d_1 的信息量为

$$\begin{aligned} \text{Entropy}(d_1) = & \frac{m_1 + n_1 + h_1 + l_1}{m + n + h + l} \text{Info}(D_1) + \\ & \frac{m_2 + n_2 + h_2 + l_2}{m + n + h + l} \text{Info}(D_2) + \\ & \frac{m_3 + n_3 + h_3 + l_3}{m + n + h + l} \text{Info}(D_3) = \\ & \frac{4}{13} \left(- \frac{2}{4} \text{lb} \frac{2}{4} - \frac{1}{4} \text{lb} \frac{1}{4} - \frac{1}{4} \text{lb} \frac{1}{4} \right) + \\ & \frac{5}{13} \left(- \frac{1}{5} \text{lb} \frac{1}{5} - \frac{2}{5} \text{lb} \frac{2}{5} - \frac{1}{5} \text{lb} \frac{1}{5} - \frac{1}{5} \text{lb} \frac{1}{5} \right) + \\ & \frac{4}{13} \left(- \frac{1}{4} \text{lb} \frac{1}{4} - \frac{1}{4} \text{lb} \frac{1}{4} - \frac{2}{4} \text{lb} \frac{2}{4} \right) = 1.662 \end{aligned}$$

故, d_1 的信息增益是

$$\begin{aligned} G(d_1) &= \text{Info}(D) - \text{Entropy}(d_1) = \\ & 1.988 - 1.662 = 0.326 \\ \text{同理可得} \\ G(d_2) &= \text{Info}(D) - \text{Entropy}(d_2) = \\ & 1.988 - 1.238 = 0.750 \\ G(d_3) &= \text{Info}(D) - \text{Entropy}(d_3) = \\ & 1.988 - 1.488 = 0.5 \\ G(d_4) &= \text{Info}(D) - \text{Entropy}(d_4) = \\ & 1.988 - 0.739 = 1.249 \end{aligned}$$

因此,根节点选择“品牌忠诚度 d_4 ”,并以此类推建立子树的节点,建立拓展分析树如图 2 所示。

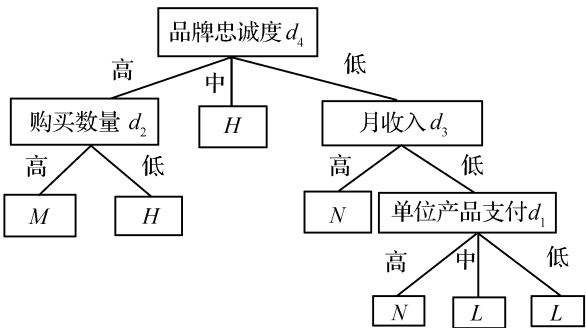


图 2 本体知识拓展分析树

Fig.2 Ontology knowledge expand analysis tree

当公司需要提高“次价值客户群(N)”的客户价值时,通过调用 `FineTreeNode(Tree T, Node H)`,得到“品牌忠诚度—月收入—单位产品支付— N ”和“品牌

忠诚度—月收入— N ” 2 条路径,调用 CoreProblem (N)函数得到影响“次价值客户群(N)”的核心问题是“品牌忠诚度”,因此,可以对“品牌忠诚度”采取可拓变换,生成提高“品牌忠诚度”策略供企业管理者选择。同理,当公司需要提高“低价值客户群(L)”的客户价值时,产生提高“品牌忠诚度”、“单位产品支付”和更换“月收入”客户群等策略;当公司需要对“潜价值客户群(H)”提高客户价值时,产生提高“购买数量”的策略供企业管理者选择。

此外,还可以通过本体知识拓展分析树对变换后进行预测,在本例子中,为提高“潜价值客户群(H)”,通过对“购买数量”进行可拓变换,可以调用预测函数 Forecast(String[] data,Node node)做定性分析,发现当“购买数量”变为“高”时,“潜价值客户”将质变为“价值客户”。

3.2 实验结果与分析

一般使用 Java 技术实现可拓策略生成系统^[11-12]。本实验也通过 Java 技术编码实现,得到如图 2 的本体知识拓展分析树,并且可以得到从根节点到任何叶子节点的路径。

在实验中,由于每个客户的当前价值由购买数量、单位产品支付和月收入 3 个属性所决定,并且各个属性只占 1/3 的比例,在不考虑传导变换的情况下,仅对其个别属性进行数值变换。

在这里,由于“低价值客户(L)”的核心矛盾是“品牌忠诚度”、“单位产品支付”和“月收入”,不具有对比性;“次价值客户群(N)”的核心矛盾是“品牌忠诚度”,这符合客户价值理论。因此,选用“潜价值客户群(H)”进行分析,通过函数 CoreProblem(H)的调用,得到“潜价值客户”的核心问题是“购买数量”,通过对潜价值客户群体的“购买数量”和“单位产品支付”这 2 个属性进行可拓变换对比分析,得到实验结果如图 3、4 所示。

可拓变换分析

可拓变换的对象为	客户ID	核心问题为	可拓变换的属性为	变换因子为	变换的结果为	变换的次数为
潜价值客户	客户4	购买数量	quantity	加1	价值客户	1 次
潜价值客户	客户6	购买数量	quantity	加1	价值客户	2 次
潜价值客户	客户7	购买数量	quantity	加1	价值客户	2 次

图 3 “购买数量”加 1 的变换结果

可拓变换分析

可拓变换的对象为	客户ID	核心问题为	可拓变换的属性为	变换因子为	变换的结果为	变换的次数为
潜价值客户	客户4	购买数量	unitpayment	1.05 倍	价值客户	1 次
潜价值客户	客户6	购买数量	unitpayment	1.05 倍	价值客户	6 次
潜价值客户	客户7	购买数量	unitpayment	1.05 倍	价值客户	7 次

图 4 “单位产品支付”1.05 倍的变换结果

Fig.4 The transform results of “unitpayment” 1.05 times

从实验结果来看,当对“潜价值客户”的“购买数量”属性进行加 1 因子可拓变换时,客户 4、客户 6、客户 7 分别由原来的 2、3、5 增加到 3、5 和 7 时就发生了质变,变换次数分别是 1 次、2 次和 2 次;当对“单位产品支付”属性进行 1.05 倍因子可拓变换时,客户 4、客户 6、客户 7 分别由原来的 450、260、220 增加到 472.52、372.88 和 325.04 才发生质变,其变换次数分别是 1 次、6 次和 7 次。从企业的角度来看,提高低价值客户的“购买数量”更具有实践性和效益性。这意味着,通过实验获得了一个可拓知识:改变潜价值客户的“购买数量”属性,比改变其“单位产品支付”属性能更快提高客户价值。

实验结果表明,通过建立本体知识拓展分析树,有效地找到了矛盾的核心问题,使得下一步有目的地对有关客户属性进行可拓变换,可以减少变换的次数,改变了以往可拓策略生成系统中通过从下(叶)到上(根)试探性的变换模式,提高了可拓策略生成效率。

4 结束语

本文结合可拓学和决策树 2 个不同领域的知识,提出了本体知识拓展分析树建立的方法,重点论述了本体知识拓展分析树在可拓策略生成系统的作用。通过分析本体知识拓展分析树得到矛盾问题的核心问题,使系统更有效地生成可拓策略解决矛盾。以提高客户价值为例进行实验,其结果说明了所提方法的有效性,使得对基元的可拓变换有依据,生成更客观、更合理的策略,同时也提高了系统的智能化水平。

将本体知识拓展分析树应用到可拓策略生成系统中,还需要进行大量的研究。比如,本体知识拓展分析树的剪枝和算法优化,如何对大量样本集数据进行处理,以及建立领域本体知识拓展分析树等,这些都将是下一步研究的重要内容。

Fig.3 The transform results of “quantity” plus 1

参考文献:

- [1] 杨春燕, 蔡文. 可拓工程[M]. 北京: 科学出版社, 2007: 1-344.
- [2] 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: 1-231.
- [3] HAN Jiawei, KAMBER M. 数据挖掘概念与技术[M]. 范明, 孟小峰. 译. 北京: 机械工业出版社, 2008: 188-200.
- [4] 孟飞翔, 帅立国, 姜昌金. 决策树在客户价值分析中的应用[J]. 计算机技术与发展, 2007, 17(4): 60-63.
- MENG Feixiang, SHUAI Liguang, JIANG Changjin. An application of decision tree to analyze the value of customer[J]. Computer Technology and Development, 2007, 17(4): 60-63.
- [5] GRUBER T R. A translation approach to portable ontology specification[J]. Knowledge Acquisition, 1995, 5(2): 199-220.
- [6] 刘宗妹. 本体可拓模型的复合元实现及应用研究[D]. 广州: 广东工业大学, 2010: 1-48.
- LIU Zongmei. Realization and application study of composite element of ontology extension model[D]. Guangzhou: Guangdong University of Technology, 2010: 1-48.
- [7] 杨春燕, 李小妹, 陈文伟, 等. 可拓数据挖掘方法及其计算机实现[M]. 广州: 广东高等教育出版社, 2010: 84-92.
- [8] 演克武, 张磊, 孙强. 决策树分类法中 ID3 算法在航空市场客户价值细分中的应用[J]. 商业研究, 2008, 371(3): 24-29.
- YAN Kewu, ZHANG Lei, SUN Qiang. The application of decision making tree classification ID3 algorithm in customer value segmentation in aviation market[J]. Commercial Research, 2008, 371(3): 24-29.
- [9] 杨兰卢, 润德. 基于客户价值的客户分类方法研究[J]. 现代管理科学, 2007(11): 95-96.
- YANG Lanlu, RUN De. Customer classification based on customer value[J]. Modern Management Science, 2007(11): 95-96.
- [10] 权明富, 齐佳音, 舒华英. 客户价值评价指标体系设计[J]. 南开管理评论, 2004, 7(3): 17-23.
- QUAN Mingfu, QI Jiayin, SHU Huaying. An evaluation index system to assess customer value[J]. Nankai Business Review, 2004, 7(3): 17-23.
- [11] 方卓君, 李卫华, 李承晓. 自助游可拓策略生成系统的研究与实现[J]. 广东工业大学学报, 2009, 26(2): 83-89.
- FANG Zhoujun, LI Weihua, LI Chengxiao. Research and realization of extension strategy generating system for independent travel[J]. Guangdong University of Technology, 2009, 26(2): 83-89.
- [12] 李承晓, 李卫华. 租房可拓策略生成系统[J]. 智能系统学报, 2011, 6(3): 272-278.
- LI Chengxiao, LI Weihua. Research on a tenement extension strategy generation system[J]. CAAI Transactions on Intelligent Systems, 2011, 6(3): 272-278.

作者简介:



温树勇, 男, 1986 年生, 硕士生研究生, 主要研究方向为智能软件。



李卫华, 女, 1957 年生, 教授, 主要研究方向为面向 Agent 计算、网络信息系统、智能软件, 发表学术论文 40 余篇。