

文章编号: 1001-0920(2010)09-1324-05

基于语义的决策模型能力评估与选择方法

张波, 向阳

(同济大学 电子与信息工程学院, 上海 201804)

摘要: 本体决策模型选择的最佳手段是使计算机在理解决策问题和决策模型自身能力的基础上进行. 通过领域本体, 决策问题和决策模型可以具备被计算机自动理解的形式化语义. 在理解决策问题语义的基础上, 系统可选择对应的求解模型类别并获取决策问题内在需求, 进而根据对应的候选模型语义对其具备的能力进行评估, 选择最适合于决策问题的决策模型. 最后, 实例分析结果表明了这种模型选择方法是有效且可行的.

关键词: 决策模型; 语义; 本体; 能力评估; 模型选择

中图分类号: TP391

文献标识码: A

Semantics based decision model capacity evaluation and selection

ZHANG Bo, XIANG Yang

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China. Correspondent: ZHANG Bo, E-mail: sh.zhangbo@gmail.com)

Abstract: The best method of decision model selection is based on fully understanding the meanings of decision problem and the capacities of decision model. Through domain ontology, decision problem and decision model are described in form of formal semantics which can be understood automatically by computer. Due to the understanding semantics of decision problem, DSS can extract the requirement of problem and select the corresponding model class which can solve the problem. Then, the capacities of candidate models are evaluated according to the semantics of decision models and eventually find out decision model which is the most complete one for the problem. Finally, the example analyzing and experiment show that the model selection method is effective and feasible.

Key words: Decision model; Semantics; Ontology; Capacity evaluation; Model selection

1 引言

决策模型是整个决策支持系统的核心组成部分. 随着当前互联网技术的飞速发展, 决策支持系统面临着如何适应网络环境、充分利用分散在网络中丰富的决策资源, 进行协作式组合决策的难题. 在传统决策支持研究中, 决策模型的构建是个重要的问题, 然而面对网络环境下存在的大量丰富决策资源, 继续自主构建模型已经成为一种负担和浪费, 决策支持系统所需要解决的关键问题是如何在丰富的网络决策资源中选择最有能力的决策模型. 传统DSS中的决策模型选择主要有两类方法: 解析法和人工智能法^[1, 2]. 解析法主要通过数学方法对决策模型进行选择, 包括: 基于目标线性规划的方法, 依靠历史信息来选择模型^[3];

基于案例统计学习的模型选择方法^[4]等.

人工智能法主要采用人工智能技术寻找决策模型, 主要包括: 基于模糊集合学习方法的模型选择^[5], 基于粗糙集方法的模型选择^[6], 基于挖掘算法的模型选择^[7], 基于神经网络的模型选择方法^[8], 基于自然语言理解的模型选择方法^[9], 面向随机过程或随机领域的模型选择^[10]等. 然而, 解析法依赖于过往数据的分析, 如果过往数据不够充分或数据存在误差, 则解析法效果会大大下降. 人工智能法无法使计算机的智能选择建立在真正理解决策问题, 以决策问题的内在含义选择适合的决策模型, 很大程度上人工智能法离不开专家指导. 模型选择是对问题进行认知并利用已有知识进行识别, 进而作出决定的过程. 模型选择最佳

收稿日期: 2009-08-04; 修回日期: 2009-11-09.

基金项目: 国家自然科学基金项目(70771077); 国家863计划项目(2008AA04Z106); 上海市科委制造业信息化专项基金项目(08DZ1122300); 国家自然科学基金青年基金项目(60903032); 上海信息化发展专项基金项目(200901015).

作者简介: 张波(1978—), 男, 江苏常州人, 博士后, 从事语义网、智能决策理论的研究; 向阳(1962—), 男, 山东泰安人, 教授, 博士生导师, 从事智能决策支持系统、人工智能等研究.

的途径应该是让计算机对决策问题进行理解, 将得到的问题含义与所拥有的模型具备的能力进行差异评估, 最终获得令决策问题满足的模型. 基于上述考虑, 本文提出了在决策问题和决策模型形式化语义表示基础上进行模型能力评估, 最终进行模型选择的方法.

2 领域本体和语义表示

定义 1 决策问题描述本体可定义为如下 6 元组:

$$DO = (C, R, OP, SP, GP, I). \quad (1)$$

其中: C 为问题的概念, 用于唯一标量某一类问题; R 为概念所拥有的与其他概念之间的所属关系集合; OP 为某一问题概念可能的初始属性集合; SP 为某一问题概念在求解过程中出现的属性状态集合; GP 用于描述问题概念可能的目标属性集合; I 为该概念的实例集合.

根据上述本体定义, 决策问题语义可表示为

$$PS = (Id|PC, S, G, GP, RelationP). \quad (2)$$

其中: 序对 $Id|PC$ 为该问题的唯一标志以及问题概念名; 非空集合 S 为该问题所拥有的描述状态集合; G 为该决策问题可能的目标集合; $RelationP$ 为指该决策问题与其他决策问题之间存在的结构关系集合.

定义 2 决策模型本体可表示为

$$MO = (MC, R, P, INPUT|OUTPUT). \quad (3)$$

其中: MC 为模型概念, R 为模型与模型的关系集合, P 为模型描述属性, 序对 $(INPUT|OUTPUT)$ 为该模型可能的输入输出状态的类型集合.

决策模型语义可表示为

$$MS = (Id|MC, MO, Input|Output). \quad (4)$$

其中: 序对 $Id|MC$ 为该模型的唯一标志以及模型概念名; 非空集合 MO 为对该模型进行描述的本体; $Input|Output$ 为该决策模型可能的输入输出变量.

3 决策模型类别选择

决策模型选择从两个方面出发: 决策问题类型和决策模型的能力. 决策模型选择时, 首先通过决策问题所属的类型来判断需选择何种对应的决策模型.

算法 1 模型类别选择算法. 该算法的具体描述如下:

```

输入: 决策问题语义;
输出: 符合类别的决策模型语义.
{
  P ← ps.PC
  For each mci ∈ MO do
  If match(ps.PC, mci) then Q ← mci

```

```

If Q ≠ ∅ then
  For each mci ∈ Q do
  If ∃msi ∧ msi.mc ∈ Q then
    return msi
  Else return error
}.

```

4 决策问题的可解决状态判定

模型选择是根据自身求解能力特点找到问题语义中的特征, 判断两者之间对应关系. 下面给出对应关系判读定义.

定义 3 对于给定的决策模型语义 ms 和决策问题语义 ps , 若存在赋值映射 N , 使得 $N_{ps.PC} \rightarrow ms.mc$, 且满足

$$\begin{aligned}
& (\exists ps.st' \in PS.ST \wedge \exists ms.input' \in \\
& MS.INPUT \wedge ps.st' = ms.input'), \\
& (\exists ps.g' \in PS.G \wedge \exists ms.output' \in \\
& MS.OUTPUT \wedge ps.g' = ms.output'),
\end{aligned}$$

则该决策问题特征项称为可被决策模型解决的.

对于给定的决策模型语义 ms 和决策问题语义 ps , 若存在赋值映射 N , 使得 $N_{ps.pc} \rightarrow ms.mc$, 且满足

$$\begin{aligned}
& (\forall ps.st' \in PS.ST \wedge \forall ms.input' \in \\
& MS.INPUT \wedge ps.st' = ms.input'), \\
& (\forall ps.g' \in PS.G \wedge \forall ms.output' \in \\
& MS.OUTPUT \wedge ps.g' = ms.output').
\end{aligned}$$

则该决策问题特征项称为可被决策模型完全解决的.

定义 4 决策问题是可被决策模型解决的, 当且仅当问题本身或其所有子问题都能找到决策模型, 并被其解决; 决策问题是可被决策模型完全解决的, 当且仅当问题本身或其所有子问题都能找到决策模型, 并被其完全解决.

算法 2 中, 函数 $find(ps.pc_i, ms.mc)$ 的作用是将问题语义特征项在决策模型语义中查找是否有与之匹配的概念, 并返回对应的模型概念名; 函数 $match(ps.st_i^j, ms.input)$ 匹配从 $find$ 函数中返回的模型概念的 $input|output$ 描述是否与决策问题语义中的初始状态集合和目标状态集合匹配.

算法 2 决策模型可解决决策问题语义判定算法. 具体算法步骤如下:

```

输入: 决策问题语义和决策模型语义;
输出: 决策问题被解决的状态.
{
  P ← ps.PC
  T ← ps.ST, G ← ps.G

```

```

For each ps.pci ∈ P do
  If find(ps.pci, ms.mc) then P ← P - {ps.pci};
  For each ps.stij ∈ T and ps.gij ∈ G do
    If match(ps.stij, ms.input)
      and match(ps.gij, ms.output)
    then T ← T - {ps.stij}, G ← G - {ps.gij}
    If P = φ and T ≠ φ and G ≠ φ
      then return understood;
    Else return error
  }.

```

5 决策模型能力的评判方法

5.1 衡量决策模型能力的指标

衡量决策模型所具备能力的标准是考量该决策模型在多大程度上能够满足决策问题. 本衡量决策模型能力指标条件可分为 3 类: 时间指标 T^g , 资源指标 R 和预期目标 G .

5.1.1 时间指标

时间指标表达了决策问题对决策模型所提出的时间要求. 时间量 $\text{time} = (\text{st}, \text{ft}, \text{lt}|\text{bt})$. 其中: st 为该任务的开始时间, ft 为该任务的结束时间, $\text{lt}|\text{bt}$ 指明执行该任务在时间点 lt 暂停执行, 并且在 bt 时间点恢复.

时间指标 $T^{g_i} = (\text{time}^{g_i}, \sigma, \theta)$, 其中: 目标 g_i 的时间量 time^{g_i} 最早开始时间为 σ , 最晚结束时间为 θ .

5.1.2 资源指标

资源指标表达在决策问题的初始条件中可向决策模型提供的辅助资源, 并指明这些资源可消耗的程度.

资源指标 $R^{g_i} = (\tilde{R}, \text{DR}, \text{SR})$. 其中: \tilde{R} 为资源最大费用值, DR 为独占资源费用值, SR 为非独占资源费用值.

5.1.3 预期目标指标

预期目标指标指明决策问题所期望达到的预期决策质量效果, 也是对目标的约束描述. 预期目标指标 $G = \langle g_i, \text{right}_i \rangle$, 目标 g_i 具有的 right_i 为预期重要度. 预期目标的重要度 right_i 可通过以下两方面计算:

1) 频率权重 $P(g_i)$. 频率权重表示目标在过往的决策样本集中出现的次数所占比重, 次数越多重要度越大. 假设 g_i 是决策问题描述本体 DO 中概念的属性, 即 $g_i \in \text{DO.pc.G}$, 则对于 DO.pc 的所有实例集合而言, 其中共出现子目标的次数为 $|\text{DO.pc.I}_j.G|$, 子目标在其所属问题类别中占有的出现频率权重为

$$P(g_i) = \frac{\sum |g_i|}{\sum |\text{DO.pc.I}_j.G|}. \quad (5)$$

2) 本体权重 $O(g_i | \text{DO})$. 本体权重体现目标在决策问题描述本体中所占的重要度. 本体权重分为概念

权重和属性权重. 假设 g_i 为决策问题描述本体 DO 中一个概念的属性, 即 $g_i \in \text{DO.pc.G}$, 本体中对应的概念为 DO.pc_i , 对应的概念属性为 $\text{DO.pc}_i.g_l$, 则有

$$O(g_i | \text{DO}) = \lambda r_{\text{pc}_i} + (1 - \lambda) r_{g_l}. \quad (6)$$

其中: r_{pc_i} 为概念 DO.pc_i 的权重; r_{g_l} 为描述概念 DO.pc_i 的 $\text{DO.pc}_i.g_l$ 子目标权重; $0 \leq \lambda \leq 1$ 为预设参数. 综合两方面内容, 预期目标的重要度为

$$\text{right}_i = P(g_i) O(g_i | \text{DO}). \quad (7)$$

5.2 决策模型能力评判计算

1) 模型响应时间能力评判

响应时间为决策模型对预期目标集从求解开始到求解结束所需要的时间.

假设决策问题所赋予的最大求解时间费用值为 \tilde{T} , 决策模型对某预期子目标 g_i 所需要执行的时间为 t^i , 子目标预期实行时间与规定的预期时间匹配函数为 $\text{match}(g_i)$, 则预期响应时间效益函数为

$$\text{value}(\text{time}) = \tilde{T} - \sum_{i=1}^n (t^i)^\eta - \sum_{i=1}^n \text{match}(g_i). \quad (8)$$

其中: η 为预设指数, 用于调节和控制决策模型执行时可能出现的预期响应的延迟效应. 比较函数 $\text{comp}(g_i)$ 反应所有子目标预期实行时间与决策问题规定的预期时间之差, 即

$$\text{comp}(g_i) = T_{\text{predict}} - T_{\text{standard}}. \quad (9)$$

2) 资源费用效益评判

资源费用效益评判是决策模型进行求解时所需要花费资源价值的效益. 假设用户系统为该问题所准备的资源最大消费值为 \tilde{R} , 则对于决策模型响应的要求资源消费值 $\text{need}.R$, 其可获得的资源效益函数为

$$\text{value}(\text{need}.R) = \tilde{R} - \sum_{i=1}^n R_i^g - \sum_{i=1}^n \text{occupied}(g_i). \quad (10)$$

其中: R_i^g 为执行预期子目标 g_i 需要的资源开销值; 函数 $\text{occupied}(g_i)$ 为求解目标 g_i 时的独占资源开销值.

3) 预期目标效益判断

决策模型的目标效益通过其可解决的预期子目标重要度来判断. 假设决策模型预期可解决子目标表示为 \tilde{g}_i , 则预期目标效益判断可计算为

$$\text{value}(\tilde{G}) = \frac{\sum \text{right}(\tilde{g}_i)}{\sum \text{right}(g_i)}. \quad (11)$$

综合 3 方面的能力评判, 最终模型的能力可计算如下:

$$\text{value}(M) = \text{value}(\text{time}) \text{value}(R) \text{value}(\tilde{G}). \quad (12)$$

6 基于语义的决策模型选择算法

基于语义的决策模型选择算法将决策问题语义

表 1 问题实例

问题代码	概念类	问题状态		问题目标		问题结构	
Pro1	能源费用预测类问题	预测期	1月	预测能源总费用	数值型	父概念	投入目标类预测类问题
		能源消耗量	200				
		能源动力费预算	1500				
		超标容忍度	0~300				
Pro2	资金需求量预测类问题	产品生产成本	18	需求满足	布尔值	父概念	投入目标类预测类问题
		资金流动周转天数	5				
		资金需求量	36000				
		超额容忍度	0~10000				
Pro3	产品销售收入分析类问题	期初产品结存量	2000	销售收入	数值型	父概念	投入目标类预测类问题
		产品销售量	3500~4000				
		产品生产量	2800~3200				
		产品出厂单价	30~32	生产情况	布尔型		
		实际成交价格	52~56				
		收益预期	>30000				
Pro4	销售分析类问题	销售产品品种号	1-50	销售总数	数值型	父概念	投入目标类预测类问题
		销售产品数量	3000~4500				

与决策模型语义进行比较, 根据候选决策模型的能力评估选择能力最强的决策模型. 本文采用如下基于语义的决策模型选择算法:

算法 3 基于语义的决策模型选择算法. 具体步骤如下:

Step1: 对于决策问题语义, 利用算法 1 判定该决策问题可被模型库中若干模型解决的程度, 分为如下 3 类情况:

Step1.1: 若存在决策模型可完全解决该决策问题, 则将这些决策模型语义定义为候选模型, 记为集合, 转 Step2;

Step1.2: 若仅存在决策模型可部分解决该决策问题, 则将这些决策模型存入, 转 Step3;

Step1.3: 若不存在决策模型可解决该问题语义, 则转 Step5.

Step2: 对集合中的候选模型进行预期效益评判, 选择效益最佳的决策模型进行问题求解, 算法结束.

Step3: 对集合中的模型进行效益评判, 选择效益最佳的决策模型. 利用这些决策模型部分求解该决策问题中的目标.

Step4: 对问题语义进行剥离操作, 分离出不可求解的目标部分, 形成新决策问题语义, 并进行语义迭代转换(基于语义迭代的目标转换方法已于另文阐述, 在此不再赘述), 形成新问题语义. 返回Step1, 若决策问题可被若干决策模型解决, 则算法退出; 若决策问题最终有部分目标不可被解决, 则转Step5.

Step5: 对问题语义进行等效目标转换. 返回Step1, 重新进行模型选择.

Step6: 提交不可解决目标语义寻求协作求解, 算法退出.

7 实例分析

下面给出一个企业效益决策分析的例子. 表 1 给出了 4 个问题语义实例; 表 2 给出了 9 个决策模型实例; 按算法 1 进行模型类别选择后, 可选择决策模型如表 3 所示; 4 个决策问题可获得候选模型解决状况如表 4 所示; 表 5 给出了决策模型能力评估时的相关属性值; 经过算法 3 的选择, 在实例中的 4 个问题分别可选择的决策模型如表 7 所示.

本文采用自主开发决策模型选择的原型系统检验本文提出的模型选择算法的有效性. 该原型系统主要进行企业营销数据分析决策, 系统模型库中具有预测、计划、分析 3 类决策模型. 模型库中所有模型均由

表 2 模型实例

模型	概念名	输入	输出
M1	一元回归预测类模型	能源动力消耗量, 能源动力单价	能源消耗总费用
M2	一元回归预测类模型	能源动力消耗量, 能源动力费预算, 超标容忍度	是否超预算, 可接受程度
M3	时间序列预测类模型	产品种类, 产品生产耗时, 总时间量	最佳效益产品生产安排
M4	回归预测模型类模型	产品生产成本, 资金流动周转天数, 资金可获得额, 资金需求量, 超额容忍度	需求是否满足, 可接受程度
M5	单目标线性规划类模型	销售产品品种量, 销售产品数量	销售总数, 产品销量排序
M6	费用效益分析类模型	期初产品结存量, 产品销售量, 产品生产量	销售总数, 生产满足情况
M7	对比分析类模型	期初产品结存量, 产品生产量	生产满足情况
M8	费用效益分析类模型	期初产品结存量, 产品销售量, 产品生产量, 产品出厂单价, 实际成交价格, 收益预期	销售收入, 销售总数, 生产满足情况, 收益程度
M9	分析类模型	销售产品品种量, 销售产品数量	销售总数, 产品销量排序

表 3 候选模型列表

决策问题	问题概念	匹配模型概念	可供选择模型
Pro1	能源动力费用 预测类问题	一元回归预测模型,回 归预测模型模型	M1, M2, M4
Pro2	资金需求量预 测类问题	一元回归预测模型,回 归预测模型模型	M1, M2, M4
Pro3	产品销售收入 分析类问题	线性规划模型,费用效 益分析模型,对比分析 模型	M5, M6, M7, M8
Pro4	销售分析问题	分析模型,线性规划类 模型,费用效益分析模 型,对比分析模型	M5, M6, M7, M8, M9

表 4 决策问题被决策模型解决的状态示例

决策问题	可选择模型	可部分解 决模型	可完全解 决模型	不可解决 模型
Pro1	M1, M2, M4	M1, M2	/	M4
Pro2	M1, M2, M4	/	M4	M1, M2
Pro3	M5, M6, M7, M8	M6, M7	M8	M5, M9
Pro4	M5, M6, M7, M8, M9	/	M5, M9	M6, M7, M8

表 5 决策模型能力评判属性示例

决策问题		\tilde{T}	\tilde{R}	DR	G
Pro4		20	50	28	1
子目标	销售总数	6	20	/	/
	销量排序	9	30	/	/
子目标		t^i	need.R	DR	G
M5	销售总数	4	16	7	0.2
	销量排序	10	22	15	0.8
M9	销售总数	7	11	2	0.3
	销量排序	9	18	13	0.7

表 6 决策模型能力评判值示例

决策模型	T^g	R	G	总判定值
M5	7	-10	1	-2
M9	3	6	1	10

表 7 最终选择决策模型

决策问题	选择决策模型	问题分解1	问题分解2
Pro1	M1, M2	M1	M2
Pro2	M4	/	/
Pro3	M8	/	/
Pro4	M9	/	/

表 8 原型系统中3类决策模型选择效果分析

模 型	选择后求解成功率/%	结果满意平均得分
预测类	81.65	85.54
计划类	78.37	83.72
分析类	79.61	87.56

自主开发的企业营销决策模型本体描述,该本体中分预测类模型、计划类模型和分析类模型3大类概念.针对这3大类概念进行细化分类,共定义了39个

决策模型概念,描述概念的属性共213个.表8给出了系统长期运行后3类决策模型的平均求解成功率和满意率.在对选择的模型求解后,用户满意率均可达到83%以上(用户满意率为用户主观满意度的人工评分,取值范围为0~100,数值越高表示用户越满意).

8 结 论

决策模型是决策问题求解的执行人,模型是否与决策问题相对应、模型的求解能力等因素直接决定问题求解结果.本文使决策问题和决策模型具有计算机可理解的形式化语义,发现问题类别所对应模型,评估模型所具备的求解能力,在语义理解和能力评判的基础上选择最能胜任问题求解的决策模型.下一步的工作是将决策模型选择方法与用户偏好相结合,研究决策模型能力信任度的评价方法,使模型选择建立在用户意志和可信任的基础上.

参考文献(References)

- [1] 向阳,杨德礼.模型选择Agent研究[J].管理科学学报,2004,7(2):47-54.
(Xiang Y, Yang D L. Research of model selection agent[J]. J of Management Science in China, 2004, 7(2): 47-54.)
- [2] 王菲,黄梯云.遗传算法在基于二元决策树的模型选择中的应用[J].管理科学学报,1999,2(2):57-70.
(Wu F, Huang T Y. The application of genetic algorithm in model selection[J]. J of Management Science in China, 1999, 2(2): 57-70.)
- [3] Klein G. Developing model string for model management [J]. J of Management Information System, 1986, 3(2): 96-110.
- [4] Sirola Miki. Using conceptual decision model in a case study [J]. Lecture Notes in Artificial Intelligence, 2003, 2773(1): 126-133
- [5] Schmidt Stefan, Chang Elizabeth, Dillon Tharam, et al. Fuzzy decision support for service selection in e-business environments [C]. Proc of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making. Honolulu, 2007: 374-381
- [6] Herbert Joseph P, Yao J T. Rough set model selection for practical decision making [C]. Proc the 4th Int Conf on Fuzzy Systems and Knowledge Discovery. Haikou, 2007: 203-207
- [7] Namin F Samimi, Shahriar K, Ataee-pour M, et al. A new model for mining method selection of mineral deposit based on fuzzy decision making [J]. J of the South African Institute of Mining and Metallurgy, 2008, 108(7): 385-395
- [8] West David, Dellana Scott. Model selection strategies for ensemble solutions to bankruptcy detection [C]. Proc of the 2005 Int Conf on Artificial Intelligence. Las Vegas, 2005: 46-52

(下转第1332页)