

基于本体的决策问题语义理解及精炼方法

张 波^{1,2}, 向 阳², 黄震华²

(1. 上海师范大学信息与机电工程学院, 上海 200234; 2. 同济大学电信学院, 上海 201804)

摘 要: 理解复杂决策问题的关键在于使决策问题所包含的领域、目标、状态、结构等信息被系统所“读懂”. 本文利用本体表达语义, 实现了问题语义关系表示和语义计算. 在保持理解结果不变的原则上, 本文提出利用语义迭代方法将计算机无法理解的问题语义转化为可完全理解的语义. 本文给出了语义精炼方法并形成最优闭合问题空间, 在结构复杂度和内容复杂度两方面具有最低复杂度. 实验分析表明, 本文提出的复杂决策问题语义理解方法是有效的.

关键词: 本体; 语义理解; 语义迭代; 语义精炼; 最优闭合问题空间

中图分类号: TP302.7 **文献标识码:** A **文章编号:** 0372-2112 (2012) 08-1603-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2012.08.017

Ontology Based Decision Problem Semantic Analysis and Refining

ZHANG Bo^{1,2}, XIANG Yang², HUANG Zhen-hua²

(1. College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 200234, China;

2. College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: A key challenge of complicated decision problem understanding is to make the information of decision problem, i. e. domain, goal, status and structure, readable and understandable for decision support system. This paper introduces ontology to describe semantics, and then problem semantic relationship and semantic computation are presented. Based on the rule of keeping the understanding results unchanged, semantic iteration method is addressed in order to transform the decision problem semantics, which cannot be analyzed through ontology, into fully understandable problem semantic. Further, semantic refining method is presented so that each complicated decision problem space can be refined into a most optimized closed problem space with a minimum complexity, which can be calculated from two facets: structure complexity and content complexity of problem space. Experiments show that the method of complicated decision problem semantic analysis is effective and feasible.

Key words: ontology; semantics understanding; semantics iteration; semantics refining; most optimized closed problem space

1 引言

所谓“有所问有所答”, 决策支持系统(DSS)必须准确理解问题才能防止“答非所问”并提供高质量决策. 然而, DSS不具备思维能力, 无法理解复杂决策问题, 这已成为制约 DSS 发展的瓶颈问题之一.

“决策问题”是由初始状态及其目标等信息组成的决策需求, 其关键复杂性反映在: 问题领域及其目标、结构、子问题状态^[1]. 本文决策问题理解研究集中在使 DSS“读懂”上述导致复杂性的因素. 从人类思维角度出发, 问题理解的最佳介质是问题中所凝结的语义. 对 DSS 而言, 问题语义应具备良好的组织结构和推理能力, 同时不带有冗余和重叠内容. 传统 DSS 所面临的问题较简单, 这种简单形式提问已无法满足当前复杂决策的需求. 因此, 复杂问题决策主要面临如下困境: (1)内

容不确定^[3]. 不确定决策方法^[4,5]难以保障问题理解准确性; (2)关键词知识理解^[6]始终停留在理解的“表面”层次, 缺乏领域知识表示的规范性和共享性; (3)缺乏自动逻辑推理的能力; (4)消除复杂度和冗余度能力不足.

本文提出了一种基于本体的复杂决策问题语义理解方法, 该方法主要具有以下特点: (1)借助本体描述决策问题形式化语义, 建立问题理解的逻辑推理基础; (2)利用语义映射使决策问题获得 DSS 理解; (3)提出语义计算、迭代等的方法, 使不被理解的问题转换为可理解的问题语义; (4)提出语义精炼方法以降低决策问题的复杂度和冗余度.

2 相关研究工作

决策问题理解已经有很多相关研究^[7-9], 这些方法虽使决策问题具有了可理解基础, 但却存在着语义推理

计算能力不足、理解后复杂度和冗余度增加等问题。

近年来,基于语义表示方法成为问题理解研究的热点.Martín 等提出了基于预先定义的知识模型进行语义表示^[10];Stanojevic 等提出层次语义表方法进行语义知识表示和处理^[11].Robert 等提出利用语义决策表工具进行问题语义进化过程处理^[12];Argent 提出基于语义模型的问题描述语言与问题概念化软件环境^[13].还有问答系统中所讨论的问题理解^[14].但问答系统仅针对简单问题,对于复杂化问题理解能力尚显不足。

如前所述,决策问题理解需要使问题成为系统“可读”的表示方式,同时实现问题从“不可读”到“可读”的推理转化.本体作为一种形式化的知识表示形式为决策问题语义的自动理解提供了有力的支持,同时目前众多研究在基于本体的决策问题理解方面做出了贡献^[15,16].但这些研究重点限于利用本体进行领域知识表达,将本体语义用于系统计算及优化进而辅助决策问题理解的研究尚处于起步阶段.本文重点研究计算基础上的决策问题形式化语义理解及效果优化问题。

3 基于本体的问题语义理解

3.1 决策问题描述本体

本文首先给出如下概念准备:

(1)决策问题描述本体 DO 是一个四元组: $DO = (PC, SR, IR, P)$,其中 PC 表示决策问题概念名; SR 表示概念之间父-子关系; IR 表示非父子关系; P 表示描述概念的属性.本体中概念权重 τ_{pc} 与属性权重 τ_p 满足如下条件*:①根概念节点的权重为 1;②存在 n 个概念 pc_i 与概念 pc' 具备如下关系: $pc_i = \{x \mid \exists x \in DO. PC \wedge SR(pc', x)\}$,则满足 $\tau_{pc'} = \sum_{i=1}^n \tau_{pc_i}$;③概念 pc_i 存在 m 个描述属性 $pc_i. p_j$,则概念所有属性的权重满足

$$\sum_{j=1}^m \tau_{p_j} = 1.$$

(2)概念 pc_1 若存在 $pc_i \in DO. PC$ 满足 $pc_i = \{x \mid SR(pc_1, x) \vee SR(x, pc_1) \vee IR(x, pc_1) \vee IR(pc_1, x)\}$,则称 pc_i 为 pc_1 的语义扩展,概念语义扩展集合记为 $\Omega(pc_1)$.

(3)若存在元素 x ,满足 x 在语义级别上与 DO 中的元素 y 一致,则称存在语义映射 N ,使 x 被 y 语义理解,记为 $N_x \rightarrow y$.

3.2 决策问题语义表示

复杂决策问题往往由若干子问题组成,而子问题又可再分为子问题.通过问题分解后子问题进行决策问题求解,被证明是有效的简化途径^[2,17].本文以此为基础,称子问题集合为决策问题空间 $\mathfrak{S} = \{ps_1, ps_2, \dots\}$,其中每个 ps_k 代表子问题语义;决策问题语义 $PS = (Id \mid C, S, G, RelationP)$.其中序对 $Id \mid C$ 为该问题的唯一

标志和对应的 $DO. PC$;非空集合 S 表示该问题所拥有的描述状态集合; G 表示该决策问题可能的目标集合; $RelationP$ 是指该决策问题与其他决策问题之间存在的结构关系集合。

3.3 决策问题语义理解算法

本文决策问题语义理解的基本思想是:利用映射 N 在 ps_k 和 DO 之间进行匹配,从而获得理解.下面给出问题语义理解的定义:

定义 1 对于给定 DO 及 ps_k ,如果存在语义映射 N ,满足 $N_{ps.c} \rightarrow DO.pc$ 和 $\exists s \in ps.S$ 并且 $N_s \rightarrow DO.pc.p$,则称 ps_k 可被 DO 理解,记为 $ps \approx DO$;如果存在语义映射 N ,满足 $N_{ps.c} \rightarrow DO.pc$ 和 $\forall s \in ps.S$ 均有 $N_s \rightarrow DO.pc.p$,且 $\forall r \in Relationship$ 均有 $(N_r \rightarrow DO.sr) \vee (N_r \rightarrow DO.ir)$,则称 ps_k 可被 DO 完全理解,记为 $ps \vdash DO$.

后文中将满足可完全理解的决策问题记为 FP ,可部分理解的问题记为 PP ,不可理解的问题记为 UP .

利用 DO 进行问题语义理解算法如下:

算法 1 决策问题理解判定算法 (evaluate)

输入:决策问题语义 ps

输出:决策问题理解状态

```
[1]  $P \leftarrow ps.c$ ;
[2]  $T \leftarrow ps.S$ ;
[3] If  $find(ps.c, DO)$  Then  $P = \emptyset$ ;
[4] For each  $ps.s \in T$  do
[5]   If  $match(ps.s_i, DO.pc.P)$  then  $T \leftarrow T - \{ps.s_i\}$ ;
[6]   If  $P = \emptyset$  and  $T = \emptyset$  then Return  $FP$ ;
[7]   Else If  $P = \emptyset$  and  $T \neq \emptyset$  Return  $PP$ ;
[8]   Else Return  $UP$ ;
[9] End
```

上述算法中,函数 $find(ps.c, DO)$ 判断是否 $N_{ps.c} \rightarrow DO.pc$ 成立,并返回布尔值与对应的 DO 概念名;函数 $match(ps.s, DO)$ 判断从 $find$ 函数返回的概念属性是否与 $ps.s$ 匹配。

4 决策问题的语义关系与语义计算

4.1 问题语义关系

决策问题结果空间 $\Gamma = (\xi_1, \xi_2, \dots)$ 是一组描述决策问题获得事实、数据、状态等类型的变量.本文将

* 本文所采用的概念权重计算方式:若是的直接子概念,则;属性复杂权重计算为,若概念存在个描述属性,则;若概念是多个不同概念的直接子概念,则概念的权重可计算为其在不同父概念处获得权重和的平均值。

$\Psi(DO) = \{ps \mid \forall ps \models DO\}$ 记为 FP 集合; Γ 是 $\Psi(DO)$ 的结果集合, 若满足 $\forall x \in \Gamma \wedge \exists y \in DO$ 均有 $N_x \rightarrow y$, 则称结果集合 Γ 为可理解的, 此时将结果集合记为 $\Gamma(DO)$.

首先定义本体与资源的组织方式: 本体-资源库.

定义 2 本体-资源库 $OR = (DO, R, con)$, 其中 DO 是决策问题描述本体; R 是决策资源描述集合, 其中存在的资源个体称为元素 ξ_i ; 二元关系 $con(\xi, x)$ 语义为 DO 中的个体 x 可对资源 ξ 进行解释.

我们将 $con(\xi, x)$ 记为 $con(\xi, DO)$, 即 DO 中存在个体 x 可以对 ξ 解释. 问题-结果对 $\vartheta = (PS', R')$ 是决策问题与决策结果组成的二元组, 且分别满足条件 $PS' \in \Psi(DO)$ 和 $R' \subseteq R \wedge R = \{\xi \mid \forall \xi \in \Gamma(DO) \wedge con(\xi, DO)\}$.

问题-结果对具有两种关系: 覆盖和等效.

定义 3 对于 ϑ_1 和 ϑ_2 , 若满足 $R'_1 \subseteq R'_2$, 称 ϑ_1 被 ϑ_2 覆盖, 记为 $\vartheta_2 \triangleright \vartheta_1$; 问题 ps_1 和 ps_2 是目标等效的, 当且仅当它们对应的 ϑ_1 和 ϑ_2 满足 $R'_1 = R'_2 \wedge ps_1 \cdot c = ps_2 \cdot c$, 记为 $ps_1 \equiv ps_2$.

目标等效问题集合称为目标等效簇.

定义 4 若存在空间 $K = \{\vartheta_1, \vartheta_2, \dots\}$, 且 $\vartheta_i = (PS'_i, R'_i)$, 则 $R = (PS^R, R^R)$ 为 K 上的目标等效簇, 当且仅当同时满足下列条件:

- (1) 对任意 $\vartheta_i, \vartheta_j \in R$, 均有 $\vartheta_i \cdot ps \equiv \vartheta_j \cdot ps$;
- (2) 对任意 $\vartheta_i, \vartheta_j \in R$, 均有 $\xi_i \in R^R \rightarrow \xi_j \in R^R$;
- (3) 对任意 $\vartheta_i \in R$, K 中不存在 $\vartheta_k \in K$ 满足 $\vartheta_k \cdot ps \equiv \vartheta_i \cdot ps_i$.

4.2 问题语义计算

为了实现问题语义的可计算性, 本文定义了三类语义计算算子: 合并, 选并和剥离.

定义 5 对 ps_1 和 ps_2 , 语义计算包括如下三类:

- (1) 合并算子 (∞) 可定义如下:

$$ps_1 \infty ps_2 = (only(id_1, id_2) \mid parent(c_1, c_2), \\ (ps_1 \cdot S \cup ps_2 \cdot S), (ps_1 \cdot G \cup ps_2 \cdot G), \\ (ps_1 \cdot RelationP \cup ps_2 \cdot RelationP))$$

其中 $only(id_1, id_2)$ 表示任取 id_1, id_2 中一个; $parent(c_1, c_2)$ 表示在本体中 c_1, c_2 的共同祖先概念.

- (2) 选并算子 (∞) 可定义如下:

$$ps_1 \infty ps_2 = (only(id_1, id_2) \mid parent(c_1, c_2), \\ (ps_1 \cdot S \cup ps_2 \cdot S), (ps_1 \cdot G \cap ps_2 \cdot G), \\ (ps_1 \cdot RelationP \cup ps_2 \cdot RelationP)).$$

(3) 剥离算子 (∇). 对于 ps_1 和 ps_2 , 且存在 $\vartheta_2 \triangleright \vartheta_1$, 则 ps_2 中剥离 ps_1 可定义如下:

$$ps_2 \nabla ps_1 = (ps_2 \cdot id \mid ps_2 \cdot c, ps_2 \cdot S, \\ (ps_2 \cdot G - ps_1 \cdot G), ps_2 \cdot RelationP)$$

对于语义计算, 有如下定理(证明略)

定理 1 问题语义计算结果相对于原问题语义, 除目标语义外, 无其他语义损耗.

5 面向决策结果的问题语义迭代

语义迭代的目的是从 UP 到 FP 转化中保持结果等效的或最近似迭代. 本节针对无目标等效簇进行的语义迭代.

5.1 近似语义迭代

本文参考文献[18]最大等价再生式和最小等价再生式思想, 将本体映射改进并扩展至语义迭代计算中.

定义 6 设问题-结果对为 $\vartheta = (PS', R')$, 本体-资源库为 $OR = (DO, R, con)$, 则:

- (1) ϑ 的最小上界:

$$Lub(\vartheta, OR) = \{\vartheta' \mid \forall \xi' \in \Gamma(DO) \wedge \exists x \in DO \\ \rightarrow con(\xi', x) \wedge \vartheta \triangleright \vartheta' \wedge (\neg \exists \vartheta'' \rightarrow \vartheta' \triangleright \vartheta'' \triangleright \vartheta)\}$$

- (2) ϑ 的最大下界:

$$Glb(\vartheta, OR) = \{\vartheta' \mid \forall \xi' \in \Gamma(DO) \wedge \exists x \in DO \rightarrow \\ con(\xi', x) \wedge \vartheta \triangleright \vartheta' \wedge (\neg \exists \vartheta'' \rightarrow \vartheta \triangleright \vartheta'' \triangleright \vartheta')\}$$

定义 7 对于问题-结果对 $\vartheta = (PS', R')$, 本体-资源库为 $OR = (DO, R, con)$, 则 ϑ 有:

- (1) 最大等价再生式: $Ler(\vartheta) = \vartheta'_1 \vee \vartheta'_2 \vee \dots \vee \vartheta'_i$,

其中 $Glb(\vartheta, OR) = (\vartheta'_1, \vartheta'_2, \dots, \vartheta'_i)$;

- (2) 最小等价再生式: $Ser(\vartheta) = \vartheta'_1 \wedge \vartheta'_2 \wedge \dots \wedge \vartheta'_i$, 其中

$Lub(\vartheta, OR) = (\vartheta'_1, \vartheta'_2, \dots, \vartheta'_i)$.

$Ler(\vartheta)$ 语义为: $ps^{Ler(\vartheta)} = \infty_{i=1}^n ps^{\vartheta'_i}$, $ps^{\vartheta'_i}$ 为 $Glb(\vartheta, OR) = (\vartheta'_1, \vartheta'_2, \dots, \vartheta'_i)$ 中 ϑ'_i 的语义;

$Ser(\vartheta)$ 语义为: $ps^{Ser(\vartheta)} = \infty_{i=1}^n ps^{\vartheta'_i}$, $ps^{\vartheta'_i}$ 为 $Lub(\vartheta, OR) = (\vartheta'_1, \vartheta'_2, \dots, \vartheta'_i)$ 中 ϑ'_i 的语义.

定理 2 对于 ϑ , 除了目标等效簇 $\mathfrak{R}(\vartheta)$, 不存在其他 ϑ' 的结果比序对 $Ler(\vartheta) \mid Ser(\vartheta)$ 的结果更接近于 ϑ 的结果.(证明略)

5.2 面向决策结果的语义迭代算法

本节首先给出问题语义迭代算法:

算法 2 语义迭代算法(substitute)

输入: 可理解决策问题语义 ps

输出: 迭代后可完全理解决策问题语义 ps^{eq}

- [1] $T \leftarrow ps \cdot S$; $S \leftarrow \emptyset$;
- [2] $Q \leftarrow ps \cdot G$; $M \leftarrow \emptyset$; $N \leftarrow \emptyset$;
- [3] For each $ps \cdot s_i \in T$ do
- [4] If $match(ps \cdot s_i, DO \cdot pc \cdot P)$ then $T \leftarrow T - \{ps \cdot s_i\}$;
- [5] else $S \leftarrow ps \cdot s_i$, $T \leftarrow T - \{ps \cdot s_i\}$;
- [6] For each $ps \cdot g_j \in Q$ do

```

[7] If  $match(ps.g_j, DO)$  then  $Q \leftarrow Q - \{ps.g_j\}$ ;
[8]   else  $M \leftarrow ps.g_j, Q \leftarrow Q - \{ps.g_j\}$ ;
[9]    $ps' = (ps.id \mid c, S, M, RelationP)$ ;
[10]   $ps'' = (ps.id \mid c, T, Q, RelationP)$ ;
[11]   $N \leftarrow convert(M)$ ;
[12] If  $find(R(ps.id^R, L)) \wedge L \subseteq N$ 
      then  $select\_max(ps.id, R), N \leftarrow N - L$ ;
[13]    $ps' \leftarrow ps^R$ ;
[14]   else  $\vartheta' = (ps', N)$ ;
[15]  $select\_max(ps'_1 \mid ps'_2, Ser(\vartheta') \mid Ler(\vartheta'))$ ;
[16]    $ps' = ps'_1 \infty ps'_2$ 
[17]  $ps^{eq} = ps' \infty ps''$ 
[16] End

```

上述算法中, $exit(\mathfrak{R}(ps.id^R, N))$ 函数测试是否存在形如 $\mathfrak{R}(ps.id^R, N)$ 的目标等效簇, 若存在则返回该等效簇; $convert(M)$ 函数用以表示将不可描述目标转换为可理解的目标集合; $select_max(ps.id, \mathfrak{R})$ 函数从 \mathfrak{R} 中选择类别与 $ps.id$ 相同的 FP 实例, 且 $\sum \tau_{ps.s}$ 值最大; $select_max(ps'_1 \mid ps'_2, Ser(\vartheta') \mid Ler(\vartheta'))$ 则表示从 $Ser(\vartheta') \mid Ler(\vartheta')$ 中选择对实例 $ps'_1 \mid ps'_2$, 且满足 $\sum \tau_{ps'_1.s}$ 和 $\sum \tau_{ps'_2.s}$ 值最大。

迭代方法正确性可通过一下定理表明。

定理 3 问题 ps 为 PP 状态, 满足对于 $\forall s \in ps.S$ 均有 $\neg \exists N_s \rightarrow DO.pc.p$, 且对于 $\forall \xi \in ps.G$ 均有 $con(\xi, DO)$, 则 ps 必可通过 DO 的问题类找到包含其预期结果的问题。

定理 4 PP 可迭代为目标等效的 FP 。

定理 5 语义迭代算法无语义损耗。

6 决策问题的复杂性度量与语义精炼

6.1 决策问题的复杂性度量方法

本文从决策问题间的结构关系和描述程度两个方面度量复杂度, 称之为结构复杂度和内容复杂度。对于 $\mathfrak{S} = \{ps_1, ps_2, \dots\}$, 其结构复杂度和内容复杂度可以分别表示为:

(1) 结构复杂度:

$$com_s(\mathfrak{S}) = \sum \tau_{pc_{class}(ps_i)} / \left[\sum \tau_{pc_{class}(ps_i)} + \sum_{|P|} \sum \tau_{pc_{class}(\Omega(ps_i))} \right]$$

(2) 内容复杂度:

$$com_c(\mathfrak{S}) = \frac{1}{n} \times \sum_{|\mathfrak{S}|} \left\{ \frac{1}{2} \times \left[\sum \tau_{Ass(ps_i.S, class(ps_i).P)} + \sum |Unass((ps_i.S, class(ps_i).P))| / |ps_i.S| \right] \right\}$$

其中 $class(ps_i)$ 函数指明 ps_i 通过 $N_x \rightarrow y$ 在 DO 对应的概

念; $|P|$ 函数表示空间的问题个数。 $Ass(ps_i.S, class(ps_i).P)$ 返回 $ps_i.S$ 中与 DO 中 $class(ps_i).P$ 相同的元素; $Unass((ps_i.S, class(ps_i).P))$ 返回不能被 DO 中 $class(ps_i).P$ 描述的属性。

对于一个复杂决策问题的复杂度由上述两个方面组成, 可以表示为:

$$com(\mathfrak{S}) = com_s(\mathfrak{S}) \times com_c(\mathfrak{S})$$

6.2 决策问题语义精炼方法

最优闭合问题空间 Φ 可定义如下:

定义 8 对于不存在目标等效簇的决策问题空间 \mathfrak{S} , 其对应结果空间为 Γ , 那么最优闭合问题空间 Φ 是一组问题个体 ps_i 组成的集合, 且其中所有 ps_i 满足下列条件:

(1) 对于 $\forall ps_i \in \Phi$, 均有存在 ϑ_i 且满足 $\Gamma = \bigcup_{i=1}^n R'_i$;

(2) 对于 $\forall ps \in \Phi$, 均有 $ps \vdash DO$;

(3) 对于 $\forall ps_i \in \Phi$, 均存在 ϑ_i 且满足 $\forall \xi \in R'_i \rightarrow \xi \in OR$;

(4) 对于 $\forall ps_i, ps_j \in \Phi$, 均存在 θ_i, ϑ_j 且满足 $(R'_i \cap R'_j = \emptyset)$;

(5) 对于 $\forall ps_i, ps_j \in \Phi$, 均不存在关系满足 $SR(ps_i, ps_j)$ 。

生成最优闭合问题空间的算法如下:

算法 3 生成闭合最有问题空间

输入: 决策问题空间 \mathfrak{S} , 结果空间为 Γ

输出: 最优闭合问题空间 Φ

```

[1] For each  $ps_i \in \mathfrak{S}$  do
[2]    $evaluate(ps_i)$ ;
[3]   If  $full\_understand(ps_i)$  then  $\Phi \leftarrow \{ps_i\}$ ;
[4]   Else  $\Phi \leftarrow \{substitute(ps_i)\}$ ;
[5] For each  $ps_j \in \Phi$  do
[6]    $\vartheta_j = (ps_j, R_j)$ ;
[7] If  $exit(\mathfrak{R})$  then  $ps^m = select\_max(ps.id, \mathfrak{R})$ ;
[8]    $\Phi \leftarrow \Phi - \{\mathfrak{R}\}$ ;
[9]    $\Phi \leftarrow \Phi \cup \{ps^m\}$ ;
[10]  If  $\vartheta'_j \triangleright \vartheta''_j$  then  $\Phi \leftarrow \Phi - \{ps''_j\}$ ;
[11]  If  $R'_j \cap R''_j \neq \emptyset$ 
      then  $ps^m = ps'_j \infty ps''_j$ ;
[12]   $\Phi \leftarrow \Phi - \{ps'_j, ps''_j\}$ ;
[13]   $ps'_j = ps'_j \nabla ps^m$ ;  $ps''_j = ps''_j \nabla ps^m$ ;
[14]   $\Phi \leftarrow \Phi \cup \{ps'_j, ps''_j, ps^m\}$ ;
[15] Return  $\Phi$ 
[16] END

```

最优闭合问题空间有如下定理:

定理 6 最优闭合问题空间复杂度最优.

定理 7 Φ 生成算法没有带来语义损耗.

$$\begin{cases} \text{迭代正确比例} = \frac{\text{iteration_ok}}{\text{iteration}} \\ \text{迭代错误比例} = \frac{\text{iteration_false}}{\text{iteration}} \end{cases}$$

7 实验分析

本节通过实验验证本文方法的有效性.实验方案如下:测试平台为基于客户语义的营销策略决策支持系统;决策问题描述本体为“客户营销策略本体”,具有决策问题实例约 230 个,营销策略实例约 370 个.

7.1 决策问题语义理解效果分析

实验的决策问题理解阶段主要分为三个组方法,包括一个实验组和两个对比组参照组 baseline1 和 baseline2:(1)实验组:采用本文提出理解方法;(2)baseline1:采用用户关键词匹配方法;(3)baseline2:仅用领域本体理解而无语义迭代方法转换.决策反馈的结果为营销案例,每个问题均包含 3~5 个子目标,共针对 300 个决策问题进行案例寻找.

我们记录每次返回的营销案例数情况:case_ok 表示可接受案例数,case_false 表示不可接受案例数,case_pre 表示案例库中应返回的最符合要求案例数(该数值人工计算得到).我们定义了验证参数 ratio₁:

$$\text{ratio}_1 = \frac{\text{case_ok}}{\text{case_pre}} \times \frac{\text{case_ok}}{\text{case_ok} + \text{case_false}}$$

图 1 给出了每进行 10 次决策后满意率的平均值,

即

$$\text{平均满意率} = \frac{\sum_{i=1}^{i+9} \text{ratio}_1^i}{10}$$

经过 300 次测试,baseline1 返回的结果获得的总体平均满意率约为 0.645,baseline2 满意率为 0.707.实验组返回的结果总体平均满意率约为 0.724.实验证明采用本文问题语义理解方法后,决策效果有明显的改善.

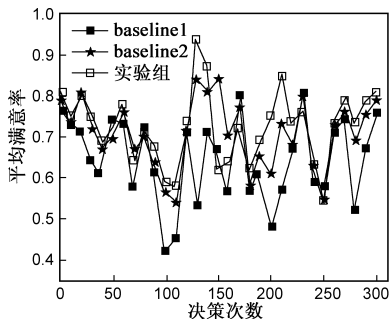


图1 决策问题语义理解效果

7.2 决策问题语义迭代效果测试

本实验重点针对语义迭代展开效果分析.我们将每一个问题进行语义迭代后产生的正确语义数量记为 iteration_ok,错误语义数量记为 iteration_false.针对语义迭代后所产生的问题语义变化定义了两个验证参数:正确比例和错误比例,计算方法如下:

实验在 50 次问题输入中加入随机的无关噪音.统计得到的迭代后获得两个验证参数.图 2 中,语义迭代正确比例可保持在 0.62 左右,而迭代后发生语义错误的噪音比例仅为 0.18.可见,语义迭代方法是有效的.

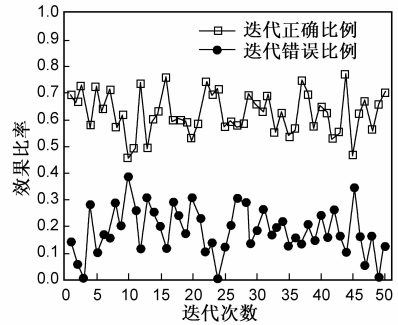


图2 问题语义迭代效果

7.3 最优闭合问题空间的复杂度精炼分析

本实验验证语义精炼方法获得最优闭合问题空间降低问题语义复杂度情况.我们将实验分为两组:(1)baseline3:原始决策问题复杂度;(2)实验组:语义精炼方法生成的最优闭合问题空间的复杂度.

本实验共进行 300 次,我们对比了两组问题的复杂度变化.图 3 中显示了最优闭合问题空间可以较明显地降低决策空间的复杂度.此外,语义精炼过程中,偶尔出现提升复杂度的情况,我们分析可能是由于语义迭代得到的问题语义比原始问题复杂所导致的.

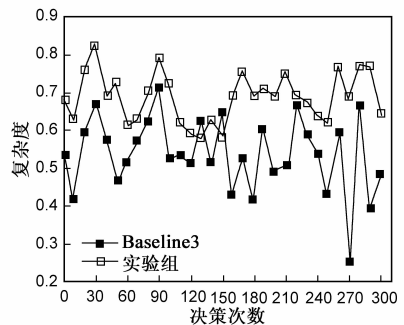


图3 问题空间语义精炼对复杂度的影响

7.4 问题理解与语义精炼方法时间影响测试

本实验旨在验证本文方法是否会带来巨大开销.实验分为两组:(1)Baseline4:本文方法系统平台;(2)实验组:安装了问题理解与语义精炼的系统平台.实验监测了两组不同系统 20 天运行,收集了 20 天内共计执行的 500 次决策的时间数据.每次决策任务在两个对比系统上均同时进行.图 4 中给出了决策从开始到结束的平均时间变化情况.实验组平均比 baseline4 多用 6.3 秒,约增加了 30% 的时间开销,这是完全可以接受的.

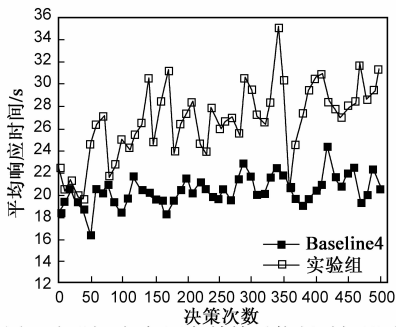


图4 问题理解与语义精炼对执行时间影响

8 结束语

日益复杂的决策问题促使 DSS 需要更有效的问题理解手段,以保证高质量决策.本文以本体为表达语义的出发点,提出了以语义计算为基础的问题理解方法.针对决策问题中不可被本体理解部分,本文采用面向决策结果的语义迭代方法,利用目标等效簇和近似等价再生式实现问题语义的有效转化,继而提出了针对复杂决策问题的最优闭合空间语义精炼方法,降低决策问题复杂度.我们下一步的工作将主要集中在语义迭代过程中如何准确找到可保持原始问题语义的替代问题,减少误差.同时,提升问题理解和语义精炼效率,减少附加系统开销.

参考文献

- [1] 向阳,于长锐.复杂决策问题求解的定性定量综合集成方法[J].管理科学学报,2001,4(2):25-31.
- [2] 于长锐,罗艳,徐福缘.复杂决策问题的多元化模型体系研究[J].管理科学学报,2004,7(2):88-94.
- [3] Smith. Managerial problem identification [J]. Omega, 1989, 17(1):27-36.
- [4] Malyshev, Piyavsky, Piyavsky. A decision making method under conditions of diversity of means of reducing uncertainty [J]. Journal of Computer and Systems Sciences International, 2010, 49(1):44-58.
- [5] Li Huaxiong, Zhou Xianzhong. Risk decision making based on decision-theoretic rough set: A three-way view decision model [J]. International Journal of Computational Intelligence Systems, 2011, 4(1):1-11.
- [6] Conractor F J, Ra W. How knowledge attributes influence alliance governance choices: A theory development note [J]. Journal of International Management, 2002, 8:11-27.
- [7] T Liang. A logical reasoning and case-bases learning in model management system [J]. Decision Support System, 1993, 10:137-160.
- [8] R. Krishnan. PDM: A Knowledge-based tool for model construction [J]. Decision Support Systems, 1991, 7:301-314.
- [9] Gangopadhyay A. Conceptual modeling from natural language

functional specifications [J]. Artificial Intelligence in Engineering, 2001, 15(2):207-218.

- [10] Martín-Bautista M J, Martínez-Folgozo S, Vila M A. A new semantic representation for short texts [J]. LNCS, 2008, 5182:347-356.
- [11] Stanojevic M, Vraneš S. Semantic approach to knowledge processing [J]. WSEAS Transactions on Information Science and Applications, 2008, 5(6):913-922.
- [12] Tang Y, Meersman R. Use semantic decision tables to improve meaning evolution support systems [J]. LNCS, 2008, 5061:169-186.
- [13] Argent. An overview of model integration for environmental applications components, frameworks and semantics [J]. Environmental Modelling and Software Volume, 2004, 19(3):219-234.
- [14] 高明霞,刘椿年.基于约束的自然语言问题到 OWL 的语义映射方法研究 [J]. 电子学报, 2007, 35(8):1598-1602. Gao Ming-xia, Liu Chun-nian. A constraints based semantic mapping method from natural language questions to OWL [J]. Acat Electronica Sinica, 2007, 35(8):1598-1602. (in Chinese)
- [15] Lee, Seok-Won, Muthurajan Divya, Gandhi Robin A. Building decision support problem domain ontology from natural language requirements for software assurance [J]. International Journal of Software Engineering and Knowledge Engineering, 2006, 16(6):851-884.
- [16] Jason J. Jung. Towards open decision support systems based on semantic focused crawling [J]. Expert Systems with Applications, 2009, 36(2):3914-3922.
- [17] 张 钺,张 玲.问题求解理论及应用 [M].北京:清华大学出版社,1990.5-14.
- [18] J Akahani, K Hiramatsu, T Satoh, Approximate Query Reformulation based on hierarchical ontology mapping [A]. International Workshop on Semantic Web Foundations and Application Technologies 2003 [C]. Nara, Japan, 2003.3:43-46.

作者简介



张 波 1978 年生于江苏省常州市.现为上海师范大学讲师,同济大学电信学院博士后.主要研究方向为本体论,决策支持系统等.
E-mail:06zhangbo@tongji.edu.cn

向 阳 1962 年生,教授,博士生导师,主要研究方向为决策支持系统,数据挖掘与计算智能等.

黄震华 1982 年生,副研究员,主要研究方向为语义计算,数据库,数据挖掘等.