

预警指标体系的本体建模及其应用

翟东升, 黄 焱

(北京工业大学经济与管理学院, 北京 100022)

摘 要: 借鉴多目标决策树中规范化的思想, 提出将一般指标体系结构规范化为由一个根节点和若干叶子节点构成的2层多叉树。根据领域本体的建模方法论, 建立基于描述逻辑的指标体系本体模型, 并利用本体编辑工具 Protégé3.1.1 将本体形式化, 给出了基于 RDF+OWL 的形式化表示。利用 Jena 开发包和 Agent 开发平台 Able 开发了预警原型系统, 从实验上论证了指标体系本体建模的正确性及其在预警系统中的应用可行性。

关键词: 本体; 指标体系; 预警系统

Warning Index-System Ontology Modeling and Its Application

ZHAI Dong-sheng, HUANG Yan

(Economics and Management School, Beijing University of Technology, Beijing 100022)

【Abstract】 Referring to the standardization about multi-objective decision-tree, the generic index-system model is standardized to a two-level tree which has a root and several leaves. According to the methodology about domain ontology modeling, the index-system ontology is modeled based on DL, and the ontology is formalized by Protégé3.1.1. The warning prototype system is established by Jena and Able, and it proves the feasibility about ontology introduced into warning system from the experiment.

【Key words】 ontology; index-system; warning system

1 概述

在预警系统中, 所有的预警工作都是围绕预警指标体系展开的, 指标体系不仅是预警的执行标准, 而且是指定了控制和调度子系统运行的规则。因此各子系统必须能够识别和理解指标体系, 即指标体系在语义层次上要为各个智能代理所共享和理解。对语义层次的信息交互和共享, 最大的技术瓶颈是语义异构问题。本体被称为“对某一概念化的清晰描述”^[1], 是关于存在的系统描述。本体技术通过提供统一的框架, 减少了领域组织内部在概念上和术语上的混淆, 从而使共同的理解和交流成为可能。

本文将本体技术引入预警系统中, 建立预警系统的指标体系本体, 从而为异构的应用系统提供共同的领域知识理解, 并为系统中基于内容的知识获取、互用和交流提供高质量的保证。

2 预警指标体系树规范化

任何一个预警任务都有自己的指标体系, 因此, 在建立其本体模型之前, 将形式多样的指标体系进行规范化, 找出指标体系的本质特征。

2.1 预警指标体系一般结构模型

本文以最常用的三级指标体系为讨论对象, 见图1。该模型分成指标层和数据层, 指标层由一级指标、二级指标和三级指标组成, 数据层由数据项组成。具体的说预警任务由一系列一级指标决定, 每个一级指标又由二级指标构成, 每个二级指标包含若干三级指标, 每个三级指标分别由一系列的数据项决定。在具体的应用中, 根据实际情况可以增加或者减少指标层的层数。每层指标之间是“与”的关系。

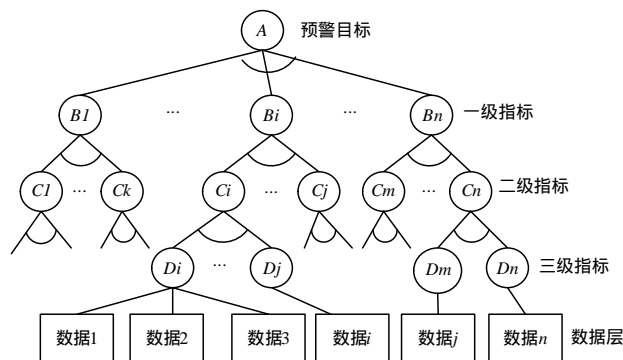


图1 指标体系模型

2.2 指标体系树的规范化

指标树规范化的基本变换步骤如下^[2]：

(1)如果指标 g 的直接分解层及其相邻层具有相同的逻辑关系, 则5层指标合并为1层, 由目标 g 直接分解层中的叶子指标和相邻层中的指标组成, 指标之间逻辑关系不变。

(2)如果指标 g 的直接分解层同时存在 AND 和 OR 逻辑关系, 则将其分解为2层, 每层指标集的指标之间逻辑关系相同。对应于逻辑关系表达式, 相当于将指标用其直接子指

基金项目: 国家自然科学基金资助重点项目(70639002); 北京市教委基金资助项目(00740); 北京市自然科学基金资助项目(9072001)

作者简介: 翟东升(1963-), 男, 副教授, 主研方向: 管理信息系统; 黄 焱, 硕士研究生

收稿日期: 2007-12-27 **E-mail:** zhaidongsheng@bjut.edu.cn

标逻辑关系表示的逻辑表达式化为析取范式，位于指标 g 直接分解层的各个子指标之间为 OR 关系，与直接分解层相邻下一层子指标集的指标之间为 AND 关系。

基于上述思想和规则，本文的指标体系树的规范化如图 2 所示。

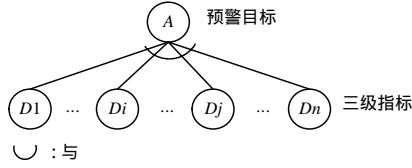


图 2 指标体系树的规范化

由图 2 看到，本系统的指标体系结构是一个多叉树，树根是预警目标，叶子节点是实际需要完成的指标项。

3 指标体系本体的术语收集与识别

3.1 类的确定

由图 2 可知，指标体系树经过规范化后是由一个根节点(预警目标)与一系列叶子节点(指标项)组成的 2 层结构的多叉树，因此，可以看到这里涉及的概念有两大类：根节点和叶子节点，分别用 *goal* 和 *leaf* 表示。用 *node* 表示节点类，*goal* 和 *leaf* 都是其子类。

3.2 属性的确定

属性包括关系属性和标注属性，关系属性标明概念之间的关系，标注属性是给概念做注释说明。

(1) 标注属性

对于根节点(预警目标)的标注属性主要是说明预警任务的总体概况。具体有：预警单位，预警计算方法，决策标准，预警完成时间，报告推送情况等。

对于叶子节点(指标项)集合如下：

$$S = \{s_1, s_2, \dots, s_i, \dots, s_n\} \quad (1)$$

其中， S 为指标集合； s_i 为第 i 个指标。

对指标 s_i 而言，其相关属性包括：由数据项计算三级指标值的数据项 $X = \{x_i | i = 1, 2, \dots, n\}$ 和其计算方法 $f(X)$ ；由三级指标计算二级指标的计算方法为 $g(X)$ ；由二级指标计算一级的计算方法为 $h(X)$ ；由一级计算预警目标的计算方法为 $k(X)$ ，即 $X \Rightarrow f(X) \Rightarrow g(X) \Rightarrow h(X) \Rightarrow k(X)$ 。

P_i 表示指标 s_i 的数据项采集点， Φ_i 表示指标 s_i 的其他约束，所以， s_i 的属性又可以表示为

$$s_i = \{P_i, X, f_i(X), g_i(X), h_i(X), k_i(X), \Phi_i\} \quad (2)$$

(2) 关系属性

所有指标项(叶子节点) S_i 是“与”的关系，即

$$S_1 \cap S_2 \cap \dots \cap S_n \Rightarrow G \quad (3)$$

用 *direct_subgoals_and* 表示。

指标集合 $S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$ 的每个指标都是互斥的，即指标项不可能相同，用 *disjoint* 表示。

指标集合中至少应有 2 个指标项，即根节点下至少要有 2 个子节点，因此对于根节点 *goal* 有至少 2 个子节点而没有父节点，在关系 *direct_subgoals_of*(直接子节点)上集的势最小为 2 即 $\min \text{Cardinality} = 2$ ，*all_subgoals_of*(所有子节点)上集的势也至少为 2；而对于 *direct_supergoals_of*(直接父节点)和 *all_supergoals_of*(所有父节点)都为 0。相反对于叶子节点只有一个父节点而没有子节点，属性约束正好相反。

用图示法表示上述术语及其关系，如图 3 所示。

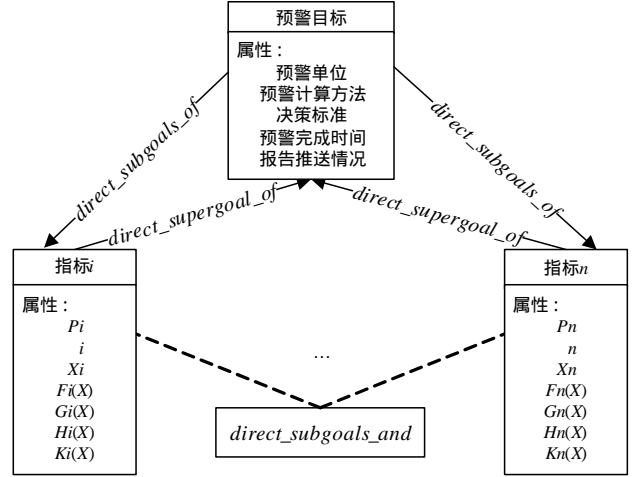


图 3 指标体系本体模型

4 指标体系本体

4.1 语义元数据模型

文献[3]提出了基于本体的元数据扩展模型，本文借鉴该文献的思想方法，并根据预警系统的实际需求，采用六元组模型： $O = \langle C, X, CD, XD, SF, CR \rangle$ 表示指标体系本体。其中， C 是指标体系术语集； X 为实例集； CD 是术语定义集； XD 为实例声明集，用来声明术语的实例； SF 为属性分配集； CR 为术语约束集合。

对于给定 $O = \langle C, X, CD, XD, SF, CR \rangle$ ，语义解释为一个 2 元组^[3]： $I = \langle \Delta', \bullet' \rangle$ 。其中， $\Delta' \neq \emptyset$ 为 O 的论域； \bullet' 是解释函数，它将 T 中的每个原子类 C 都映射为 Δ' 的一个子集 $C' \subseteq \Delta'$ ，将 T 中的每个原子属性 P 都映射为一个二元关系 $P' \subseteq \Delta' \times \Delta'$ ，将 X 中的每一个个体 a 映射为 Δ' 中的一个元素 $a' \in \Delta'$ ，语义解释将语义元数据模型与集合理论对应，从而阐明术语的语义。预警指标体系本体的语义模型如下：

$O\text{-index-system} = \langle \{node, goal, indicator, direct_supergoal_of, direct_subgoals_of, all_supergoals_of, all_subgoals_of, direct_subgoals_and, yjdw, jsff, jcbz, wcsj, tsqk, Pi, i, Fi, Gi, Hi, Ki\}, \{exp, indicator_1, \dots, indicator_n\}, \{goal \ node = 0, direct_supergoal_of \ 2, direct_subgoals_of = 0, all_supergoals_of \ 2, all_subgoals_of \ 2, direct_subgoals_and, indicator \ node = 1, direct_supergoal_of = 0, direct_subgoals_of = 1, all_supergoals_of = 0, all_subgoals_of\}, \{goal(exp), indicator(indicator_i), direct_supergoal_of(indicator_i, exp), direct_subgoals_of(exp, indicator_i), all_supergoals_of(indicator_i, exp), all_subgoals_of(exp, indicator_1, \dots, indicator_n), direct_subgoals_and(indicator_1, \dots, indicator_n)\}, \{goal \subseteq \forall yjdw.xsd: String \ \forall jsff.xsd: String \ \forall jcbz.xsd: String \ \forall wcsj.xsd: String \ \forall tsqk.xsd: String, indicator \subseteq \forall Pi.xsd: String \ \forall i.xsd: String \ \forall Fi.xsd: String \ \forall Gi.xsd: String \ \forall Hi.xsd: String \ \forall Ki.xsd: String\}, \{goal \subseteq node, indicator \subseteq node, goal \ indicator = \emptyset\} | i = 1, 2, \dots, n \rangle \quad (4)$

(1) 指标体系术语集 C 与其定义集 CD

$C = \{node, goal, indicator, direct_supergoal_of, direct_subgoals_of, all_supergoals_of, all_subgoals_of, direct_subgoals_and, yjdw, jsff, jcbz, wcsj, tsqk, Pi, i, Fi, Gi, Hi, Ki\}$

其中，*node*(节点)，*goal*(根节点-预警目标)和 *indicator*(叶子节点-指标)是原子术语，*direct_supergoal_of*，*direct_subgoals_of*，*yjdw*，*Pi* 等是属性术语。

在式(4)中， $goal \ node = 0, direct_supergoal_of$ 和 $indicator \ node = 1, direct_supergoal_of$ 都是术语定义，分别定义了根节点 *goal* 和叶子节点 *indicator*。

(2)指标实例集 X 与其声明集 XD

实例集是个体(实例)的集合。实例声明集包括类的实例声明、属性的实例声明、实例不等声明：

1)类的实例声明，记为 $C(a)$ ，表示个体 a 属于类 C 。

2)属性的实例声明，记为 $P(a,b)$ ，表示个体 a 与 b 之间存在关系 P ，并称 b 是 a 关于 P 的属性值。

3)实例不等声明，记为 $a \neq b$ ，表示个体 a 与 b 不同。

4)给定解释 I ，如果类的实例声明 $C(a)$ 成立，则有 $a' \in C^I$ 。如果属性的实例声明 $P(a,b)$ 成立，则有 $(a',b') \in P^I$ 。如果实例不等声明 $a \neq b$ 成立，则有 $a' \neq b'$ 。式(4)中 $goal(exp)$ 和 $indicator(indicator_i)$ 都是类实例声明，分别声明了 exp 是 $goal$ 的实例， $indicator_i$ 是 $indicator$ 的实例。 $direct_supergoal_of(indicator_i, exp)$ 是属性实例声明，表示 $indicator_i$ 和 exp 之间存在关系 $direct_supergoal_of$ ，并且 exp 是 $indicator_i$ 关于 $direct_supergoal_of$ 的属性值。 $direct_subgoals_and(indicator_1, indicator_2, \dots, indicator_n)$ 表示指标实例之间是“与”的关系。

(3)属性分配集 SF

属性分配集的目的就是将属性术语分配给类术语，属性分配集可表示为

$$SF = \{C_1 \subseteq D_{11} \cap D_{12} \cap \dots \cap D_{1m_1}, C_2 \subseteq D_{21} \cap D_{22} \cap \dots \cap D_{2m_2}, \dots, C_n \subseteq D_{n1} \cap D_{n2} \cap \dots \cap D_{nm_n}\}$$

其中， $C_i \in C$ 为原子类术语； D_i 表示属性类型。在式(4)中，

$goal \subseteq \forall yjdw.xsd:String \quad \forall jsff.xsd:String \quad \forall jcbz.xsd:String \quad \forall wcsj.xsd:String \quad \forall tsqk.xsd:String, indicator \subseteq \forall Pi.xsd:String \quad \forall i.xsd:String \quad \forall Fi.xsd:String \quad \forall Gi.xsd:String \quad \forall Hi.xsd:String \quad \forall Ki.xsd:String$

表示 $goal$ 这个类具有属性 $yjdw$ (预警单位)， $jsff$ (计算方法)， $jcbz$ (决策标准)， $wcsj$ (完成时间)和 $tsqk$ (推送情况)，并且这几个属性都是字符串 $String$ 类型。同理， $indicator$ 的定义类似。

(4)指标体系术语约束集 CR

术语约束集由若干个术语关系(包含关系、等价关系、非交关系)组成，用来限定 C 中术语之间的关系。 $goal \subset node$ ， $indicator \subset node$ 表示 $goal$ 是 $node$ 的子类， $indicator$ 是 $node$ 的子类， $goal \quad indicator = \emptyset$ 表示这 2 个类不存在交集。

4.2 本体形式化的实现

本体形式化的最后一步就是将抽象的语义模型用本体描述语言来表达，这里采用本体编辑工具 Protégé3.1.1 建立本体，并生成以 OWL + RDF 描述的形式化本体模型，以便机器能够读取和理解。

首先建立预警目标(根节点)和指标项(叶子节点)2 个集合的本体，如图 4 和图 5 所示。



图 4 指标体系本体

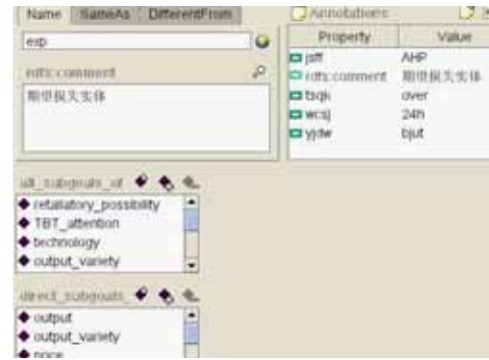


图 5 预警目标实例

由图 4 看到，对于根节点通过 $all_supergoals_of=0$ 和 $direct_supergoal_of=0$ 的属性限制保证该节点是根节点，由于指标体系至少要有 2 个指标项，因此根节点的叶子节点数量至少是 2，即 $all_subgoals_of \geq 2$ ；由于预警指标集合与指标项集合没有交集，因此两者之间的关系是 $Disjoint$ 。

以上建立的本体是指标体系的一个抽象概念模型，对于一个具体的指标体系是该模型的一个具体实例。这里以中国农产品技术贸易壁垒预警指标体系为实例，根据前文介绍的指标体系树规范化，预警目标为期望损失 exp ，将中间层的指标抛去，剩下叶子节点层 $f1, f2, \dots, f9$ ： $f1 = consumption$ ； $f2 = output$ ； $f3 = price$ ； $f4 = technology$ ； $f5 = price_variety$ ； $f6 = output_variety$ ； $f7 = retaliatory_possibility$ ； $f8 = TBT_attention$ ； $f9 = difference$ 。

预警目标是“ exp ”(期望损失)，这里将它作为“ $goal$ ”类的一个实例，如图 5 所示。

图 5 中“ exp ”的属性“ $all_subgoals_of$ (所有子结点)”的值是“ $retaliatory_possibility$ (报复可能性)”等，即表示“ exp ”的“所有子目标”是“报复可能性”等 9 个指标。同样属性“ $direct_subgoals_of$ ”标明了“ exp ”的“直接子目标”也是“ $price$ (对象国同类产品价格)”等 9 个指标。通过 $jsff$ 、 $tsdk$ 、 $wcsj$ 和 $yjdw$ 是标注属性来对预警目标实体“ exp ”进行注释。同理，“ exp ”的 9 个预警指标也是 9 个实例。

最后生成以 RDF + OWL 为本体描述语言的文档，OWL 实现类结构的表示，RDF 实现对属性的表示。这里限于篇幅，不列出文档，在 RDF+OWL 文档中，可以很方便地对本体库进行增加、删除和修改，实现对本体库(知识库)的更新。

5 预警指标体系本体的应用

以 Java 为开发平台，通过调用本体开发包 Jena 和 Agent 开发包 Able 的相关 API，开发了预警系统的原型系统，系统结构如图 6 所示。

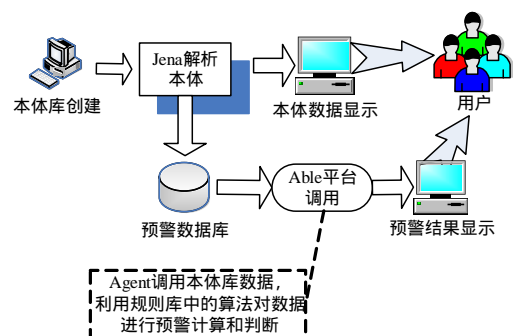


图 6 原型系统框架

(下转第 269 页)