

# 基于类自然语言的管理问题理解方法研究

卢涛, 李一军, 黄梯云

(哈尔滨工业大学管理学院, 哈尔滨 150001)

**摘要:**提出一种可用于模型管理系统的具有类自然语言特点的问题描述语言——PDL语言,给出了PDL语言的基本构造方法,指出基于该语言的问题理解过程包括问题的语义理解和问题的段落理解两个阶段,着重研究了将用PDL语言描述的问题转换为问题结构表示的问题理解方法。

**关键词:**模型管理; 问题理解; 模型构建

**中图分类号:** C931

**文献标识码:** A

**文章编号:** 1007-9807(2003)01-0037-07

## 0 引言

模型在管理决策中起着重要的作用。模型管理系统是一个对模型的建立、存储、运行和维护等过程进行管理的软件系统,在这一类系统中,模型和模型的求解算法一般是彼此独立的。模型建立后,系统调用相应的求解程序得到问题的解。对于用户来说,需要把求解的问题抽象为数学模型输入到系统中,由系统负责模型的求解。目前的模型管理系统提供多种建立模型的语言<sup>[1~5]</sup>,如代数式语言(AMPL)、结构化建模(SML,MDL)、图文法和逻辑建模(PM)等。对于非专家用户来说,直接用系统提供的语言构建问题的模型有很大的难度,这一点妨碍了用户在决策过程中对于模型的使用。

对于大多数决策者来讲,一个实用的模型管理系统,应该允许用户用自然语言描述问题,由系统自动建立模型并求解。将自然语言处理技术用于模型管理目前已有一些研究成果,如文献[6]在将自然语言句子进行切分和语法分析之后,根据中心词查找语料库,建立相应的求解模型。

由于自然语言的复杂性,自然语言处理技术无法做到对完全自然语言的理解。为此,本文提出

一种可用于模型管理系统的具有类自然语言特点的问题描述语言——PDL语言,它的优点是用户易于使用,同时因为没有二义性,系统处理的准确度高。在给出PDL语言的基本构成后,提出一种独立于模型的问题结构表示方法,以及基于PDL语言的将问题描述转化为问题结构表示的问题理解方法,与类似的工作相比,这种理解方法强调知识的应用。

## 1 PDL语言的基本构成

自然语言难以处理不仅因为它的词汇量大,而且因为每个词的词性、词义、词与词之间的联系等在不同的上下文环境中的变化规律难以把握,使得系统在处理时无法做到准确的理解。PDL语言是一种类自然语言,它以接近自然语言的形式描述问题,同时对词的使用方法、词义和句子的语法结构加以限制,使系统易于做到准确的理解。

PDL语言根据问题领域的不同拥有不同的词汇集合,但对词的定义方式是一致的。这里没有采用传统的说明词性的方法定义词汇,而是根据应用的需要来说明每个词的词义。

描述一个求解的问题,主要是说明问题中的

各种数据.系统内部可以采用数据分析中普遍应用的 E-R 图的方法描述数据之间的关系.如图 1 是一个生产计划领域涉及的部分数据之间关系的 E-R 图.

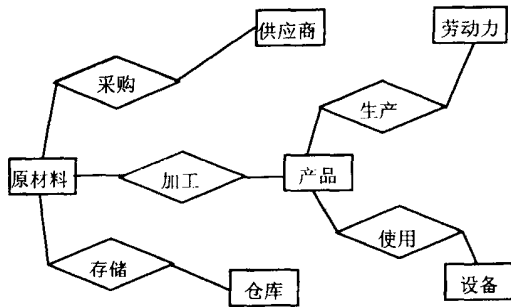


图1 生产计划领域数据 E-R 图

**定义 1** “实体”、“联系”、“属性”这 3 个词称为系统的元词汇,即系统用来定义其它词汇的词汇.

通过系统的元词汇可以定义实体类、联系类或属性类的词汇.如“产品”这个词表示一种实体,它具有“成本”、“价格”、“产量”等属性;“采购”这个词表示“供应商”和“原材料”两个实体之间的一种多对多联系,它具有“供货时间”、“数量”等属性;“价格”这个词是具有属性值的属性类,可以是产品、原材料等实体的属性.

**定义 2** 由系统元词汇定义的词汇,称为系统的领域元词汇,可以定义领域的词汇如“冰激凌”可以用“产品”来定义,它具有产品类的属性.

这样系统的一部分表示实体的词汇可以通过树型结构组织起来,其中每个结点是其父结点的子类或实例.

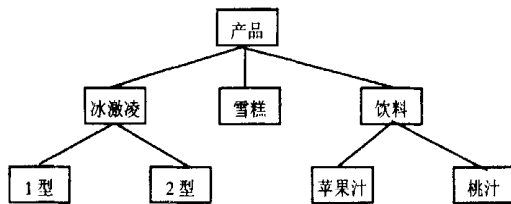


图2 实体类词汇构成的树型结构

在描述问题时通常还需要另外一类词汇,如“总产量”、“设备利用率”等.这一类词表示数据量,通常对应在某些实体属性上的计算,把这一类词汇称为计算型词汇,对这一类词汇词义的定义需要描述计算过程.

**定义 3** 领域元词汇、由领域元词汇定义的

词汇及其计算型词汇构成了问题描述语言的基本词汇集.

在基本词汇集的基础上,加上有限个辅助性词汇,就可以描述一个求解的问题.PDL 语言采用的语法也是相对简单的,可以用 BNF 范式进行定义,如:

- <句子> ::= “查询” <数据项>
- <句子> ::= <数据项> / “的值”  
为 <实数>
- <数据项> ::= <实体类词汇> <属性> /  
<计算型词汇>

事实上,句子的形式往往决定句子的语义类型,如上例中第一个 BNF 决定这一类句子的语义是确定所求的数据项.在实际系统中,可以通过输入界面的选项、按钮等控件控制输入的语法结构.

根据上面的构成规则,可以在不同的问题领域定义 PDL 语言的词汇和语法结构,PDL 语言主要是说明问题领域中包含的数据和数据关系,因此适于管理领域中定量问题的描述.

## 2 问题结构的表示

在模型管理系统中,用户输入问题的描述,系统应该根据问题自动建立求解的模型.因为模型的形式、求解目标等各不相同,所以有不同的构模方法.系统在对输入的问题描述进行处理时,应该将其转化为一种独立于模型的中间表示,称为问题的结构表示.

一个好的问题表示形式应该面向问题,同时易于转化为模型.可采用类似结构模型<sup>[3,7]</sup>的表示方法.

**定义 4** 问题域中的元素分为 5 类:原子元素,合成元素,属性元素,函数元素和测试元素.

- (1) 问题域中某类实体的一个实例,称为一个原子元素.
- (2) 问题域中两类实体的实例  $E_1, E_2$  之间的一个联系  $R$  称为合成元素,  $E_1, E_2$  称为  $R$  的定义元素.
- (3) 原子元素或合成元素  $E$  的一个数值属性称为  $E$  的属性元素
- (4) 函数元素可递归定义如下:
  - 1) 若  $e_1, e_2, \dots, e_n$  是属性元素,  $f$  是一个数值

型的  $n$  元函数,若  $e_{n+1} = f(e_1, e_2, \dots, e_n)$ , 则  $e_{n+1}$  是一个由  $e_1, e_2, \dots, e_n$  定义的函数元素  $e_1, e_2, \dots, e_n$  称为  $e_{n+1}$  的定义元素.

2) 若  $e_1, e_2, \dots, e_n$  是属性元素或函数元素,  $f$  是一个数值型的  $n$  元函数,若  $e_{n+1} = f(e_1, e_2, \dots, e_n)$ , 则  $e_{n+1}$  是一个由  $e_1, e_2, \dots, e_n$  定义的函数元素.  $e_1, e_2, \dots, e_n$  称为  $e_{n+1}$  的定义元素.

(5) 若  $e_1, e_2, \dots, e_m$  是属性元素或函数元素,  $c_1, c_2, \dots, c_k$  是常数,  $m + k = n, p$  是一个  $n$  元谓词, 若  $e_{n+1} = p(e_1, e_2, \dots, e_m, c_1, c_2, \dots, c_k)$  则  $e_{n+1}$  是一个由  $e_1, e_2, \dots, e_m, c_1, c_2, \dots, c_k$  定义的测试元素.  $e_1, e_2, \dots, e_m, c_1, c_2, \dots, c_k$  称为  $e_{n+1}$  的定义元素.

在某些情况下,一个元素可以是属性元素,同时又是函数元素,如某种产品的总成本是其单位成本与产量的乘积,总成本是属性元素,同时又是由单位成本和产量两个属性元素定义的函数元素.

一个问题域中所包含的元素和元素之间的关系可用有向图的形式表示.

**定义 5** 一个问题结构图是一个无回路的有向图  $G = (V, E, A)$ , 其中:  $V$  是有向图中顶点的集合, 每个顶点表示问题域的一个元素;  $E$  是定义有向边的集合, 表示元素之间的定义关系, 若  $(v_1, v_2) \in E$ , 即元素  $v_1$  有一条指向元素  $v_2$  的有向边, 当且仅当  $v_1$  是  $v_2$  的一个定义元素;  $A$  是属性有向边的集合,  $(v_1, v_2) \in A$ , 当且仅当  $v_1$  是原子元素或合成元素,  $v_2$  是  $v_1$  的属性元素.

如图 3 是一个简单的问题结构图, 方框表示实体元素或合成元素, 圈表示属性元素、函数元素或测试元素, 实线表示定义边, 虚线表示属性边.

问题结构图可以说明问题中各个元素的组成结构, 对于一个可求解的问题, 还应该指明哪些属性值是已知的, 哪些属性值是需要求解的.

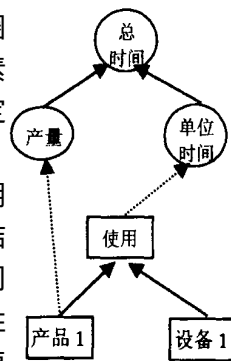


图 3 问题结构图示例

**定义 6** 一个问题的结构化表示是一个 3 元组  $(G, K, D)$ . 其中:  $G$  是

问题结构图;  $K$  是已知量集合, 即其值是由用户直接输入的或可以从数据库中直接获取的;  $D$  是决策变量的集合.

### 3 基于 PDL 语言的问题理解方法

问题理解指的是将用户输入的描述一个问题的字符串转化为对应问题的结构化表示的过程, 包括词法分析、语法分析、语义理解和问题段落理解 4 个阶段.

理解过程首先对输入字符串进行词法分析和语法分析, 即对每个句子按词进行切分, 确定每个词的类别, 识别句子的语法结构. 因为 PDL 语言的构成比较简单, 在上述过程一般不会出现二义性, 所以应用现有的自然语言处理技术可以使这个问题得到很好的解决.

#### 3.1 语义理解

根据句子的语法结构可以判断句子所表达的语义的类型, 如“产品  $P_1$  的单位成本为 0.5”, 采用的是前面例子中第一个 BNF 范式规定的语法结构, 这一类句子的语义是说明一个已知元素. 在语义分析阶段, 语义类型决定系统的动作, 如生成问题结构图中的一个元素, 将某个量加入已知量集合等; 句子中的基本词决定系统动作的对象, 每个基本词都对应问题结构图中的元素. 实体类词汇对应一个原子元素; 联系类词汇对应一个合成元素; 属性类词汇对应一个属性元素; 计算型词汇对应一个函数元素. 如对于“产品  $P_1$  的单位成本为 0.5”, 系统对应的动作是:

(1) 如果不存在与产品  $P_1$  对应的原子元素, 则生成原子元素  $p_1$ , 表示产品  $P_1$ , 将  $p_1$  加入问题结构图中;

(2) 如果不存在与  $p_1$  的单位成本对应的原子元素  $p_1$  的属性元素, 则生成属性元素  $a_1$ , 表示  $p_1$  的单位成本, 将  $a_1$  和属性边  $(p_1, a_1)$  加入问题结构图中;

(3) 将与  $p_1$  的单位成本对应的元素  $a_1$  加入到已知元素集合, 值为 0.5.

又如“要求总利润不小于 100 万元”, 这句话系统对应的动作是:

(1) 如果不存在与总利润对应的函数元素,

则生成函数元素  $f_1$ , 表示总利润, 将  $f_1$  加入问题结构图中;

(2) 生成测试结点  $t_1$ , 对应的谓词是“ $f_1$  的值 100 万元”, 将定义边  $(f_1, t_1)$  加入到问题结构图中.

为此, 系统中需要存储一张知识表, 表中规定每种语法结构对应的动作.

PDL 语言的语义理解过程是为每个基本词, 在问题结构图中生成相应的元素, 然后完成知识表中规定的动作.

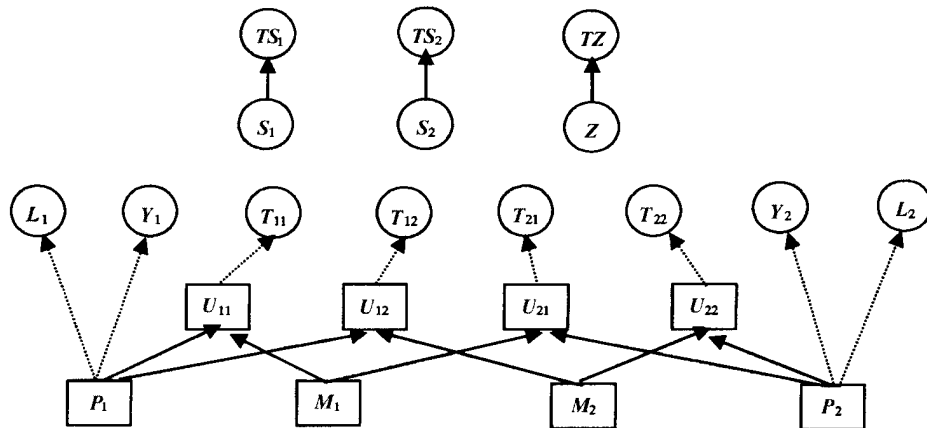
### 3.2 问题段落理解

在语义分析的基础上, 需要进一步进行段落理解. 因为描述一个问题的段落的语义不是每个句子的语义之和, 必须在整体上对问题进行理解. 具体地讲, 在语义分析阶段, 某些求解问题中需要

的元素和关系可能无法识别出来, 如在问题描述中给出某种产品的单位成本和产量, 但可能不会给出二者的关系和这种产品的总成本这个元素, 这就需要系统利用领域知识进行理解. 下面是一个产品组合问题.

“产品 1 在设备 1 上的加工时间为 0.5, 在设备 2 上的加工时间为 0.8, 产品 1 的利润为 3, 产品 2 在设备 1 上的加工时间为 0.7, 在设备 2 上的加工时间为 0.4, 产品 2 的利润为 6, 设备 1 和设备 2 每天使用不能超过 8h, 求使总利润最大, 产品 1 和产品 2 的日产量.”

图 4 是经过语义理解后的问题结构图, 从图中可以看出, 求解问题所需要的关系并没有完全反映出来, 解决这个问题, 需要应用领域的知识, 把与求解问题有关的关系识别出来.



$P_1$ :产品 1     $P_2$ :产品 2     $M_1$ :设备 1     $M_2$ :设备 2     $U_{11}$ :产品 1 在设备 1 上加工  
 $U_{12}$ :产品 1 在设备 2 上加工     $U_{21}$ :产品 2 在设备 1 上加工     $U_{22}$ :产品 2 在设备 2 上加工  
 $L_1$ :产品 1 的利润     $Y_1$ :产品 1 的产量     $L_2$ :产品 2 的利润     $Y_2$ :产品 2 的产量  
 $T_{11}$ :产品 1 在设备 1 上单位加工时间     $T_{12}$ :产品 1 在设备 2 上单位加工时间  
 $T_{21}$ :产品 2 在设备 1 上单位加工时间     $T_{22}$ :产品 2 在设备 2 上单位加工时间  
 $S_1$ :设备 1 的使用时间     $S_2$ :设备 2 的使用时间     $Z$ :总利润  
 $TS_1:S_1 = 8$      $TS_2:S_2 = 8$      $TZ:Z$ 最大

图 4 问题经过语义分析后得到的问题结构图

#### 3.2.1 系统领域知识表示

E-R 图可以表示领域数据的组成和联系, 其中实体的属性是系统能够处理的最小数据单位. 实体的数值型属性及其计算型词汇所对应的量, 统称为数据项. 相同或不同实体的数据项之间, 同一数据项的不同取值之间往往存在一些函数关系, 如产品的单位成本与产品数量的积是这一类

产品的总成本; 另外某些数据项的取值作为某类模型的输入, 经过求解后, 可得到某些数据项的值. 因此描述领域数据知识除了应用 E-R 图以外, 还应指明数据项之间的关系.

数据项的这种关系可以用框架结构描述. 一种关系对应一个框架, 框架的基本结构如图 5 示. 其中, 输入变量的个数在某些情况下是不固定的,

如根据某种产品以往的销售量预测本月的销售量,这时输入变量取值需要在构模过程确定。

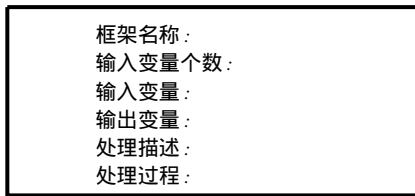


图 5 领域知识框架结构

如果一个数据项在一个框架的输入变量中出现,则称这个数据项被该框架引用,若在输出变量中出现,则称该数据项被该框架定义。系统中还应建立一个包括所有数据项的列表,为每个数据项指出引用和定义它的框架名称。

### 3.2.2 段落理解方法

段落理解过程需要对经过语义理解生成的问题结构化表示进行进一步处理,直到对问题完全理解。

**定义 7** 可求解的元素可递归定义如下:

(1) 若  $e$  是一个属性元素或函数元素,  $e \in K$  或  $e \in D$ , 即  $e$  是一个已知元素或决策变量, 则称  $e$  为可求解的元素。

(2) 若一个函数元素  $e$  的所有定义元素是可求解的, 则称  $e$  为可求解的元素。

(3) 若一个测试元素  $e$  的所有定义元素是可求解的, 则称  $e$  为可求解的元素。

**定义 8** 若一个问题的结构图中所有属性元素、函数元素和测试元素都是可求解的元素, 则称该问题是完全理解的。

段落理解过程是这样的, 首先将问题结构图中的属性元素、函数元素和测试元素拓扑排序, 按照拓扑顺序标记可求解的元素, 然后对图中每个没有定义的函数元素, 搜索领域知识框架, 在结构图中定义该元素。下面用一个递归过程 search 完成搜索。

PROC search( $v, G, G'$ )

(1) 查找数据项列表, 得定义  $v$  的领域知识框架集合  $F$ ;

(2)  $G' := G$

(3) WHILE  $v$  在  $G'$  中不是可求解的 AND  $F \neq \emptyset$  DO

    【 $f := F$  中一个框架;

$E := f$  引用的数据项的集合

        FOR 所有  $e \in E$ , 且  $e \notin V(G')$

            将  $e$  加入到图中

        flag = true

        WHILE 存在  $e \in E$  且  $e$  在  $G'$  中不是可求解的 DO

            【 search ( $e, G', G'$ );

                IF  $v$  在  $G'$  中不是可求解的

                    THEN flag = false

                    ELSE  $G' = G'$

            】

        IF flag = false

        THEN 将  $f$  从  $F$  中删除,  $G' := G$ 】

下面给出问题的段落理解过程。

输入: 问题结构图  $G = (V, E, A)$ , 已知量集合  $K$ , 决策变量集合  $D$ 。

输出: 修改后的问题结构图  $G$ 。

(1) 将  $V$  中的属性元素、函数元素和测试元素拓扑排序, 形成拓扑序列  $L$ ;

(2) 按  $L$  中元素的顺序标记可求解元素

(3) WHILE  $V$  中存在入度为 0 的函数元素  $v$  DO

    【 search ( $v, G, G'$ );

        IF  $v$  在  $G'$  中的入度  $< > 0$

        THEN  $G = G'$

        IF  $v$  的入度 = 0

        THEN 退出理解过程

    】

例 1 经过段落理解可以得到图 6 所示的问题结构图。

## 4 基于问题理解的模型管理系统体系结构

基于问题理解的模型管理系统的体系结构如图 7 所示。

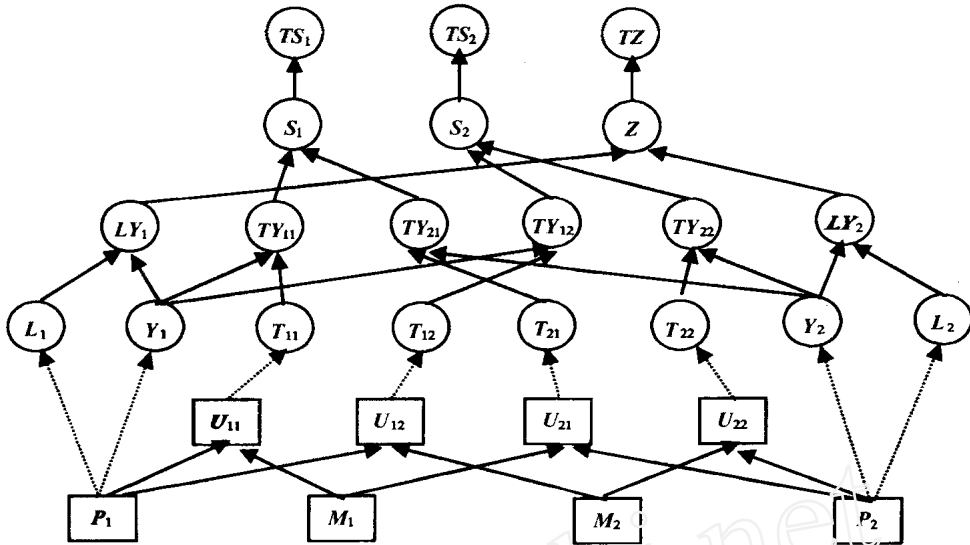
人机对话接口是用户与系统交互的接口。该模块能够接受用户输入的信息, 并输出与用户交互的内容。

问题理解模块对用户输入进行理解, 将理解的结果传送到问题求解模块; 如果无法做到对问题的完全理解, 则将结果传送到对话流管理模块。

对话流管理模块首先根据当前的问题结构, 运用领域知识, 决定对话的内容, 将对话内容按不同的优先级排列后, 输出到语言处理模块, 以用户能够理解的形式输出。

语言处理模块存储不同的输出模板, 根据对

话内容和类型选择相应的模板后,将词汇按自然语言的语序排列输出.



$P_1$ :产品1  $P_2$ :产品2  $M_1$ :设备1  $M_2$ :设备2  $U_{11}$ :产品1在设备1上加工  
 $U_{12}$ :产品1在设备2上加工  $U_{21}$ :产品2在设备1上加工  $U_{22}$ :产品2在设备2上加工  
 $L_1$ :产品1的利润  $Y_1$ :产品1的产量  $L_2$ :产品2的利润  $Y_2$ :产品2的产量  
 $T_{11}$ :产品1在设备1上单位加工时间  $T_{12}$ :产品1在设备2上单位加工时间  
 $T_{21}$ :产品2在设备1上单位加工时间  $T_{22}$ :产品2在设备2上单位加工时间  
 $TY_{11}$ :产品1使用设备1的时间  $TY_{21}$ :产品2使用设备1的时间  $LY_1$ :产品1的总利润  
 $TY_{12}$ :产品1使用设备2的时间  $TY_{22}$ :产品2使用设备2的时间  $LY_2$ :产品2的总利润  
 $S_1$ :设备1的使用时间  $S_2$ :设备2的使用时间  $Z$ :总利润  
 $TS_1:S_1 \ge 8$   $TS_2:S_2 \ge 8$   $TZ:Z$ 最大

图6 段落理解后得到的问题结构图

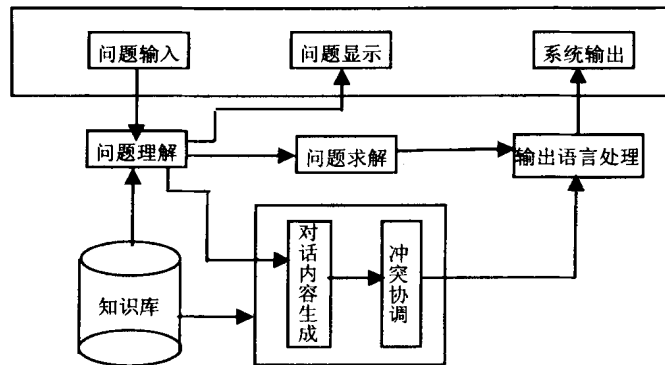


图7 基于问题理解的交互式模型构建系统的体系结构

### 5 结论

理解用户描述的问题,将其转化为系统的问题表示形式,并在此基础上构建模型是模型管理

系统的一个主要功能.为了使用户能方便地建立模型,本文提出了一种具有类自然语言特点的问题描述语言 PDL 语言,并研究了基于该语言的问题理解机制,解决了非专家用户难以有效地建立模型的问题.

## 参考文献:

- [1] Krishnan R. PDM: A knowledge-based tool for model construction[J]. *Decision Support Systems*, 1991, 7(4): 301—304
- [2] Krishnan R. A logic modeling language for automated model construction[J]. *Decision Support Systems*, 1990, 6(2): 123—152
- [3] Geoffrion A M. An introduction to structured modeling[J]. *Management Science*, 1987, 33(5): 547—588
- [4] Chari K, Sen T K. An implementation of a graph-based modeling system for structured modeling (GBMS/ SM) [J]. *Decision Support System*, 1998, 22(2): 103—120
- [5] Sen T K, Chari K. A graphical modeling system: applications in organizational model management[J]. *Omega*, 1997, 25(2): 241—253
- [6] 向阳, 黄梯云. 基于管理问题理解的智能构模理论框架研究[J]. *管理科学学报*, 1999, 2(3): 51—58
- [7] Geoffrion A M. The formal aspects of structured modeling[J]. *Operation Research*, 1989, 37(1): 30—51
- [8] Binbasioglu M. Process-based reconstructive approach to model building[J]. *Decision Support Systems*, 1994, 12(2): 97—113
- [9] 黄梯云, 李一军, 周宽久. 模型管理系统及其发展[J]. *管理科学学报*, 1998, 1(1): 57—63
- [10] Bhargava H K, Krishnan R. Computer-aided model construction[J]. *Decision Support Systems*, 1993, 9(1): 91—111
- [11] 黄梯云, 冯玉强, 周宽久. 决策支持系统中的建模知识表示研究[J]. *管理科学学报*, 2001, 4(1): 45—51
- [12] 胡祥培, 修立军, 杨德礼. 生产作业计划问题的知识表示研究[J]. *管理科学学报*, 1999, 2(2): 45—48

## Research on method of managerial problem understanding based on natural language-like problem description language

LU Tao, LI Yi-jun, HUANG Ti-yun

School of Management, Harbin Institute of Technology, Harbin 150001, China

**Abstract:** This paper present a problem description language PDL, which is natural language-like and can be applied in model management system. The structure of PDL is presented. The process of problem understanding based on PDL includes two phases: problem semantic understanding and problem paragraph understanding. This paper focuses on the method of problem understanding, by which problem description in PDL is translated into structured problem representation.

**Key words:** model management; problem understanding; model construction