

报表生成智能代理的知识表达与推理

Knowledge Representation And Reasoning In Report Generating Intelligent Agent

苏子林 (烟台鲁东大学交通学院 264025)

张洪杰 (烟台海颐软件股份有限公司 264025)

徐群参 (烟台鲁东大学计算机科学与技术学院 264025)

摘要:本文首先介绍了用于智能报表系统的知识概念;随后探讨了报表生成智能代理的集成知识表达方法;提出了报表框架匹配的评价函数和相应的推理算法;并给出了基于多代理的原型系统。

关键词:知识表示 框架 智能报表系统 软件代理

1 引言

代理技术被誉为软件开发的重大突破,在工业控制、电子商务、电子游戏和数据库应用等方面得到了广泛的研究与应用。管理信息系统(MIS)项目一般都集成智能报表系统或开发通用报表模块,因此探讨代理技术在报表自动生成过程的应用具有重要意义。文章探讨了报表生成代理的知识表示和推理算法,并给出了一个基于多代理的智能报表系统原型。

2 报表生成代理的知识概念

报表生成代理作为智能报表系统的核心,其所有功能基于报表的知识。文献^[2]引入背景图思想,将报表分为背景层和数据层,其背景层实际上是对报表格式的定义;文献^[3]提出报表知识的原始数据层、数据处理层和格式层描述。结合报表生成的规则和约束,报表知识主要包括报表格式、数据和生成三类。

2.1 报表格式知识

中国报表一般包含表格线,而且表格线先设计好,然后填入数据。报表的分区主要包括标题、表头、表体和表尾四部分^[4],其中表体区在数据分组的情况下,又可以分为组标题和组尾注区。因此报表格式可以看作由基本的单元格组成,其基本概念定义如下:

(1) 定义 1 单元格。cell = { row, col, width, height, bgcolor, forecolor, style, border, font }。其中 row、col、width、height 和 bgcolor、forecolor 分别为单元格的行号、列号、宽度、高度背景颜色和前景颜色;

样式 style = alignleft | alignright | aligncenter | alignup | alignbottom;边框 border = { leftbdr, rightbdr, topbdr, bottombdr } (border 的元素都为线 line = { width, style, color });文字 font = { content, name, size, style }, content 为文字的内容。

(2) 定义 2 标题区。caption = { firstrow, firstcol, lastrow, lastcol }。用于显示报表的标题(包括副标题),标题内容存在单元格中。

(3) 定义 3 表头区。header = { firstrow, firstcol, lastrow, lastcol, extend }。开始行(firstrow)、开始列(firstcol)、终止行(lastrow)和终止列(lastcol)的缺省值都为 0, lastcol = 0 表示缺省为最大列数;extend = 0|1,表示表头区的列数是否随着数据集内容的增加而自动扩展。

(4) 定义 4 表体区。mainbody = { firstrow, firstcol, lastrow, lastcol, extend, column }。其中 Extend = { vertical, horizontal }。表示行数和列数是否随着数据集内容增加而自动扩展;column 为报表的分栏数目,缺省为 1。

(5) 定义 5 表尾区。endnote = { rows, cols, formula }。表尾区紧接表体区显示。其公式(formula)用于定义其单元格的计算公式,公式中的行号和列号为相对编号。

2.2 报表数据知识

报表在一定格式中将数据显示出来。报表数据主要包括来源、描述及与格式的关联信息,其基本概念定义如下:

(1) 定义 6 数据源。datares = { name, dbtype, dbname, user, pwd }。用于封装与各个业务信息系统

数据库的连接信息,由数据源标识(name)、数据库类型(dbtype)、数据库名称(dbname)、账号(user)和密码(pwd)组成。

(2) 定义 7 数据集。dataset = {id, name, type, parameters, datares, content}。封装取自业务数据库的数据描述,其类型(type)有业务库表、数据视图、SQL 语句、存储过程和库外数据集等五种类型;数据集参数(parameters)指进行数据检索时需要的参数;内容(content)描述库外数据集的内容或库内数据的字典信息。

2.3 报表生成知识

报表在生成过程中,将数据在格式中按照一定规则显示。通过报表生成知识,报表数据和格式联系起来,其基本概念定义如下:

(1) 定义 8 报表框架。reprframe = {id, name, type, parameters, define, index}。其中,报表类型(type)主要包括简单报表、主从报表、分栏报表、嵌套报表、分组报表、图表报表、交叉报表和套打报表等几种;报表参数(parameters)指进行报表生成时需要的参数变量,包括变量类型和变量名称;框架定义 define = {dataset, startcol, startrow, fillstyle, parameters}是数据集的集合,定义数据集在报表格式中显示的起始行号、列号、填充方式和检索参数。index 为根据一定规则生成的索引标识,用于报表框架的定义。

(2) 定义 9 报表。reprfrm = {id, name, reprframe, parameters, user, date}。其中,报表名称(name)缺省由 reprframe. Name 和报表生成日期(date)组成;报表参数(parameters)指进行报表生成时需要的参数值;操作员(user)表示报表生成操作的用户。

3 代理知识的表达

在人工智能领域,知识的表达主要有谓词逻辑法、产生式规则、语义网络和框架等多种表示法,但是在实际应用中,单独一种表示方法难以满足要求,需要将各种表示方法集成,采用混合知识表示法^[5]。根据面向对象技术,可以将知识分为问题类,方法类,采用框架表示,可由属性槽、关系槽、规则槽等组成^[6]。

问题类知识用于描述求解问题的过程,主要有数据源定义、数据集定义、报表框架定义、报表定义、报表格式生成、报表数据获取和报表生成等。多个基本问题求解过程可以生成新的问题求解过程。其中用于报

表框架定义的一个报表框架的框架描述如下:

```
reprframe_define( //框架名称
(id(value id_value) (if_needed rpfm_id_gen)) //id 槽
(name(value null) (if_needed ask_frm_name)) //
name 槽
(type(value null) (if_needed ask_frm_value)) //
type 槽
(parameter(value para_value) (if_needed ask_frm_
para)) (define(value rpfm_def_value) (if_needed
rpfm_def)) (index(value rpfm_index) (if_needed
rpfm_index_def))
)
```

方法类知识是从一类实例或一类模型中抽象出的共性知识,它们之间可以是继承和成员的关系,也可成为对象类。在报表生成代理中主要有人机交互方法、报表框架模板确定和参数定义。人机交互方法主要包括界面框架、缺省数据和可选数据获取。

4 知识的推理过程和主要算法

在代理知识表达中,已经具有各种基本问题的求解过程定义,因此推理过程就是将问题分解,然后匹配基本问题求解过程的知识框架进行求解。知识框架中的 if_needed, if_added 或 if_deleted 等侧面是附加过程,在求解过程中直接启动,调用相应算法。

在报表框架的定义过程中, rpfm_index_def 过程在报表框架的 index. value 确定时启用。rpfm_index_def 的主要算法过程如下:

- (1) 生成报表类型代码 type_code;
- (2) 搜索框架内引用的数据集,依次生成代码 dataset_code,并用间隔符"/"连接;
- (3) 检查框架中报表格式的颜色、线型、字体、单元格尺寸、单元格样式等信息,生成代码 style_code;
- (4) 将生成的代码用间隔符"/"连接,并返回。

报表框架定义中的 rpfm_def 过程,确定报表框架引用的数据集和它们在报表格式中的分布,主要算法过程如下:

- ① 选择需要的所有数据集,得到 datasets = {set-code1, setcode2, ...};
- ② 确定报表的基本格式,如颜色、线型、字体、单元格尺寸、单元格样式等信息,生成代码 style_code;

③ 根据评价函数 func_model , 在已经定义的报表框架中, 搜索评价函数值最低的报表框架, 若评价函数的结果未超过阈值, 则得到最佳模板, 转向⑤;

④ 根据报表类型、数据集和基本格式生成缺省报表框架;

⑤ 启动报表框架界面, 显示报表框架模板, 并指导用户编辑;

⑥ 检查, 并保存报表框架; 此算法中评价函数的定义如下:

$$\text{func_model} = \text{if}(\text{type_code}_1 - \text{type_code}_0 = 0, 0, 1) * w_1 + \left(\sum_{i=1}^n \text{if}(\text{setcode}_{i1} - \text{setcode}_{i0} = 0, 0, 1) \right) * w_2 + \text{if}(\text{style_code}_1 - \text{style_code}_0 = 0, 0, 1) * w_3$$

其中 n 为报表框架中引用数据集的数量; 类型代码、数据集代码和格式代码中下标 1 和 0 分别表示新建报表框架和已有报表框架; w_1, w_2 和 w_3 分别为类型代码、数据集代码和格式代码的权重。在工程实际中, 报表类型在报表框架建立过程中, 相对于其他因素最重要; 数据集在一个报表中一般不会超过 4 个; 报表格式比重最小, 因此权重分别确定为 8, 2 和 1。在步骤③中根据报表框架的索引标识 index , 进行启发式搜索。

5 系统原型

采用上述报表知识处理方法的智能报表系统原型如图 1 所表示, 是一个多代理系统 (MAS)。考虑系统的可靠性和分布性, 采用集中与分布控制方式相结合的混合控制方式, 包括报表资源代理、报表定义代理和报表生成代理三个代理群, 以及管理代理、任务代理和协调代理。管理代理的主要任务是代理注册, 监控代理的运行, 传达全局信息和进行代理评价。协调代理的主要任务是协调代理的冲突与合作, 分配任务, 并监控任务的执行。任务代理主要处理与业务 MIS 的通信, 接受任务, 分解任务并返回任务的运行结果。报表资源代理主要进行报表资源的管理, 包括数据源、数据集、报表框架、报表的授权、备份、删除和检查功能, 以及报表数据的分析和维护功能。报表定义代理进行报表资源的定义, 包括数据源、数据集、报表框架和报表的定义。报表生成代理根据报表框架的知识描述, 解释报表知识, 提取报表数据, 并生成报表。

原型系统内采用可扩展标记语言 (xml) 进行知识

和数据的传递, 任务代理与业务 MIS 采用标准的 KQML 进行通信。系统的实现基于 J2EE 平台, 完全采用 JAVA 语言; 报表知识完全存入 SQL SERVER2000 数据库。报表生成代理在报表框架定义过程中, 直接生成 XML 文档保存报表的格式信息, 在报表生成代理进行格式知识处理时, 直接调用, 以加快输出速度。报表框架模板的搜索算法, 在 PIV2. 4G 微机上, 对于 1000 个报表框架的数据量, 搜索时间不超过一秒; 算法的阈值应随着报表框架的数据的增加逐渐调低, 范围在 1 到 4 之间。原型系统进一步证明, 系统的知识表达和推理可靠, 与传统的通用报表系统相比, 操作更加简便和灵活, 推理效率大大提高, 软件的适应性和智能性大大增强。

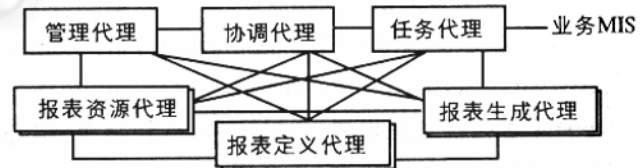


图 1 原型系统的结构

6 结束语

文章探讨了报表生成智能代理的集成知识表达方法, 提出了报表框架匹配的评价函数和相应的推理算法, 并给出了一个原型系统。原型系统采用了多代理技术, 对于此技术在智能报表系统中的应用具有重要意义。系统的进一步研究主要是分布于不同服务器的多个报表生成代理的协调和报表数据的分析。

参考文献

- 1 谢赞福, 基于 Agent 的 Web 报表自动生成通用系统 [J], 计算机工程, 2004. 6 (30)。
- 2 王元珍、汪皓, 达梦智能报表工具的设计与实现, 计算机工程与应用, 2001. 4 (37)。
- 3 高军、李晓明, 基于三层结构的电力信息系统智能报表的研究 [J], 电力科学与工程, 2003. 2 (23)。
- 4 晋玉星、马骏, 通用报表的 XML 描述 [J], 航空计算技术, 2002. 3 (33)。
- 5 宋久鹏、高国安, 混合知识表示法在基于实例设计中的应用研究 [J], 计算机工程, 2001. 11 (27)。
- 6 闫礼祥、覃征, 基于智能体的混合知识自适应推理控制 [J], 2003. 12 (24)。