



## (12) 发明专利

(10) 授权公告号 CN 102301324 B

(45) 授权公告日 2015. 08. 19

(21) 申请号 201080006214. X

(22) 申请日 2010. 01. 29

(30) 优先权数据

61/148, 888 2009. 01. 30 US

(85) PCT国际申请进入国家阶段日

2011. 08. 01

(86) PCT国际申请的申请数据

PCT/US2010/022593 2010. 01. 29

(87) PCT国际申请的公布数据

W02010/088523 EN 2010. 08. 05

(73) 专利权人 起元技术有限责任公司

地址 美国马萨诸塞州

(72) 发明人 J. 古尔德 S. 斯图德

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 邵亚丽

(51) Int. Cl.

G06F 7/00(2006. 01)

(56) 对比文件

US 2008256014 A1, 2008. 10. 16,

US 2008256014 A1, 2008. 10. 16,

US 2007050340 A1, 2007. 03. 01,

CN 1314634 A, 2001. 09. 26,

审查员 吴瑶

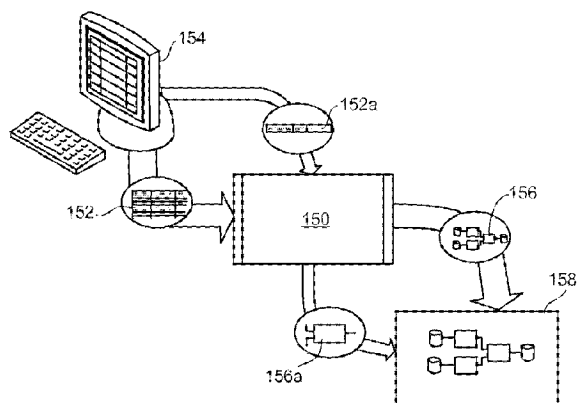
权利要求书2页 说明书12页 附图9页

(54) 发明名称

使用矢量字段处理数据

(57) 摘要

公开一种方法,包括:接收用于基于一个或多个输入值生成输出值的具有至少一个规则例(210a-h)的规则(152);生成用于从输入数据集(100)接收数据的转换(112,156),并且基于所述规则(152)转换所述数据包括生成用于输出数据集(120)中的至少一个输出变量(508,801)的第一系列的值(813),在所述第一系列的值(813)中的至少一个包括第二系列的值(816),并且提供相应于所述输出数据集(120)中的至少一个输出变量(508,801)的输出字段用于存储所述第二系列的值(816)。



1. 一种处理数据的方法,包括:

接收用于基于一个或多个输入值产生输出值的具有至少一个规则例的规则,

生成用于从输入数据集接收数据的转换,并且基于所述规则转换所述数据,其中,基于所述规则转换所述数据包括 (1) 产生用于输出数据集中的至少一个输出变量的第一系列的值、以及 (2) 产生所述第一系列的值中的至少一个值,所述至少一个值包括第二系列的值,以及

提供相应于所述输出数据集中的至少一个输出变量的输出字段用于存储所述第二系列的值。

2. 如权利要求 1 所述的方法,其中,所述转换包括在由图表示的基于图的应用的组件中,所述图中的顶点表示组件,并且所述图中的顶点之间的有向链接表示组件之间的数据的流动。

3. 如权利要求 2 所述的方法,其中,包括转换的第一图组件提供从输入数据集到所述转换的数据的流动。

4. 如权利要求 3 所述的方法,其中,第一图组件是可执行的计算组件,并且所述图包括第二图组件,该第二图组件是表示输入数据集的数据存储器组件。

5. 如权利要求 1 所述的方法,其中,产生用于输出数据集中的至少一个变量的第一系列的值包括产生用于输出表格的行,每行定义具有用于包括所述输出变量的一组变量的值的记录。

6. 如权利要求 1 所述的方法,其中,提供用于存储第二系列的值的输出字段包括提供用于存储预定数量的第二系列的值的阵列,所述预定数量是可修改为用户指定数量的默认数量。

7. 如权利要求 1 所述的方法,其中,所述输出字段包括表格中的单元格。

8. 如权利要求 1 所述的方法,其中,接收所述规则包括接收规则表格的至少一行,该行相应于规则例,并且具有包括输入值、预定值或从一个或多个输入值计算的值中的一个或多个或其组合的输出。

9. 如权利要求 8 所述的方法,其中,所述规则例包括下述中的一个或多个:具有等于阈值的输入值、具有高于阈值的输入值、具有低于阈值的输入值、具有属于一组值的输入值、具有匹配多个值的模式的输入值、具有对另一输入值的关系、具有对另一组规则的输出值的关系、或者具有对存储器中的值的关系。

10. 如权利要求 1 所述的方法,其中,所述输入数据集包括具有用于标量变量和矢量变量的值的记录。

11. 如权利要求 10 所述的方法,其中,所述记录中的至少一个包括用于存储预定数量的记录的阵列,所述预定数量是可修改为用户指定数量的默认数量。

12. 如权利要求 10 所述的方法,其中,所述记录中的至少一个包括内部参考表格以对于所述记录中的至少一个中的子记录定义关键字关系。

13. 如权利要求 12 所述的方法,还包括:响应于规则,基于所述内部参考表格中的关键字关系、对于输出数据集中的输出变量产生第二系列的值。

14. 如权利要求 1 所述的方法,还包括:响应于规则中的规则例,触发所述规则例以产生用于所述输出数据集中的输出变量的值。

15. 如权利要求 14 所述的方法, 其中, 触发所述规则例包括基于满足所述规则中的至少一个规则例的输入数据集中的标量值来触发所述规则例。

16. 如权利要求 14 所述的方法, 其中, 触发所述规则例包括基于满足所述规则中的至少一个规则例的输入数据集中的矢量中的每个值来触发所述规则例。

17. 如权利要求 14 所述的方法, 其中, 触发所述规则例包括基于应用于满足所述规则中的至少一个规则例的输入数据集中的矢量的聚集函数的输出来触发所述规则例。

18. 如权利要求 1 所述的方法, 其中, 生成所述转换包括将所述规则中的多个规则例中的每个变换为逻辑表达式以形成多个逻辑表达式, 以及将所述多个逻辑表达式编译成为计算机可执行的代码。

19. 如权利要求 18 所述的方法, 其中, 编译所述多个逻辑表达式包括组合表达式、优化单独的表达式以及优化表达式组中的一个或多个。

20. 如权利要求 14 所述的方法, 其中触发所述规则例包括由所述输入值满足预先确定的一个或多个标准。

21. 如权利要求 1 所述的方法, 还包括:

响应于所述输入值触发属于所述规则的两个或更多个规则例; 以及

通过所述转换产生两个或更多个输出值作为第二系列的值, 其中每个输出值相应于所述被触发的两个或更多个规则例。

22. 如权利要求 21 所述的方法, 还包括向用户提供检查的能力, 用于检查响应于所述输入值而被触发的两个或更多规则例。

23. 如权利要求 1 所述的方法, 其中, 所述至少一个规则例包括一个或多个标准, 当所述输入值满足所述标准时导致通过所述转换产生一个或多个输出值。

24. 如权利要求 1 所述的方法, 还包括相应于在第二系列的值上执行的操作而在所述输出数据集中存储标量值。

25. 如权利要求 1 所述的方法, 还包括当将光标放在所述输出字段之上时显示所述第二系列的值。

## 使用矢量字段处理数据

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2009 年 1 月 30 日提交的美国申请 No. 61/148,888 的优先权,通过引用的方式将其合并于此。

### 技术领域

[0003] 本说明涉及使用矢量字段处理数据。

### 背景技术

[0004] 一些计算机系统提供用于指定规则的接口,所述规则用于在各种数据处理应用中做出自动化决定。例如,与处理表示信用卡交易或航线飞行常客程序的数据相联系的决定可以由一组给定的规则来控制。在某些情况下,这些规则以人类可读形式描述。计算机系统可以提供用于用户定义或编辑这些规则的接口,然后将这些规则合并进数据处理系统。

### 发明内容

[0005] 在一个方面,概括地说,一种方法包括:接收用于基于一个或多个输入值生成输出值的具有至少一个规则例的规则;生成用于从输入数据集接收数据的转换,并且基于所述规则转换所述数据包括产生用于输出数据集中的至少一个输出变量的第一系列的值,在所述第一系列的值中的至少一个包括第二系列的值,并且提供相应于所述输出数据集中的至少一个输出变量的输出字段用于存储所述第二系列的值。

[0006] 多个方面可以包括一个或多个如下特征。

[0007] 所述转换可以包括在由图表示的基于图的应用的组件中,所述图中的顶点表示组件,并且所述图中的顶点之间的有向链接表示组件之间的数据的流动。

[0008] 包括转换的第一图组件可以提供从输入数据集到所述转换的数据的流动。

[0009] 所述第一图组件可以是可执行的计算组件,并且所述图可以包括第二图组件,该第二图组件是表示输入数据集的数据存储组件。

[0010] 产生用于输出数据集中的至少一个变量的第一系列的值可以包括产生用于输出表格的行,每行定义用于包括所述输出变量的一组变量的值的记录。

[0011] 提供用于存储第二系列的值的输出字段可以包括提供用于存储预定数量的第二系列的值的阵列,所述预定数量是可修改为用户指定数量的默认数量。所述输出字段可以包括表格中的单元格。

[0012] 接收所述规则可以包括接收规则表格的至少一行,该行相应于规则例,并且具有包括输入值、预定值或从一个或多个输入值计算的值中的一个或多个或其组合的输出。

[0013] 所述规则例可以包括下述中的一个或多个:具有等于阈值的输入值、具有高于阈值的输入值、具有低于阈值的输入值、具有属于一组值的输入值、具有匹配多个值的模式的输入值、具有对另一输入值的关系、具有对另一组规则的输出值的关系、或者具有对存储器中的值的关系。

[0014] 所述输入数据集可以包括具有用于标量变量和矢量变量的值的记录。所述记录中的至少一个可以包括用于存储预定数量的记录的阵列,所述预定数量是可修改为用户指定数量的默认数量。所述记录中的至少一个包括内部参考表格以对于所述记录中的至少一个中的子记录定义关键字关系。

[0015] 所述方法还可以包括:响应于规则,基于所述内部参考表格中的关键字关系、对于输出数据集中的输出变量产生第二系列的值。

[0016] 所述方法还可以包括:响应于规则中的规则例,触发所述规则例以产生用于所述输出数据集中的输出变量的值。触发规则例可以包括:基于满足所述规则中的至少一个规则例的输入数据集中的标量值来触发所述规则。

[0017] 触发规则例可以包括:基于满足所述规则中的至少一个规则例的输入数据集中的矢量中的每个值来触发所述规则。

[0018] 触发规则例可以包括:基于应用于满足所述规则中的至少一个规则例的输入数据集中的矢量的聚集函数的输出来触发所述规则。

[0019] 生成所述转换可以包括:将所述规则中的多个规则例中的每一个变换为逻辑表达式以形成多个逻辑表达式,以及将所述多个逻辑表达式编译成为计算机可执行的代码。

[0020] 编译所述多个逻辑表达式可以包括组合表达式、优化单独的表达式以及优化表达式组中的一个或多个。

[0021] 在另一方面,概括地说,一种计算机可读介质,存储用于更新基于图的计算中的组件的计算机程序,所述基于图的计算具有通过链接表示数据流动的元素来连接的数据处理组件,所述计算机程序包括用于导致计算机执行如下操作的指令:接收用于基于一个或多个输入值产生输出值的具有至少一个规则例的规则,生成用于从输入数据集接收数据的转换,并且基于所述规则转换所述数据包括产生用于输出数据集中的至少一个输出变量的第一系列的值,在所述第一系列的值中的至少一个包括第二系列的值,以及提供相应于所述输出数据集中的至少一个输出变量的输出字段用于存储所述第二系列的值。

[0022] 在另一方面中,一种系统包括:用于接收用于基于一个或多个输入值产生输出值的具有至少一个规则例的规则的装置,处理器,被配置为生成用于从输入数据集接收数据的转换,并且基于所述规则转换所述数据包括产生用于输出数据集中的至少一个输出变量的第一系列的值,在所述第一系列的值中的至少一个包括第二系列的值,以及用于提供相应于所述输出数据集中的至少一个输出变量的输出字段以存储所述第二系列的值的装置。

[0023] 本发明的其他特征和优点从如下描述和权利要求中将变得清楚。

## 附图说明

[0024] 图 1 是描绘示例转换的示意图。

[0025] 图 2 是示例转换生成器。

[0026] 图 3 和图 4 是示例规则集。

[0027] 图 5 是示例多激发 (Fire-Many) 规则集。

[0028] 图 6、图 7 和图 8 是示例输出、规则和结果表格片段 (tab)。

[0029] 图 9 是描绘标量和矢量的计算的示意图。

[0030] 图 10A 和图 10B 示出具有记录矢量的示例输入记录。

## 具体实施方式

[0031] 商业规则可以被表示为标准的集合,例如,所述标准可用于将数据从一种格式变换到另一种格式、做出关于数据的决定或基于输入数据的集合生成新数据。例如,在图 1 中,在航班预定系统中的记录 102 指示在航线中的乘客姓名 104、乘客在本年度已经飞行的英里 106、乘客的机票的等级 108 以及乘客的当前排 110。商业规则可以指示这样的乘客应该被分类在登机组“1”内,例如,组 118。商业规则对于人类来说通常是容易理解的,例如,“头等舱乘客在组 1”中,但是在将商业规则用于操作数据之前、可能需要将其转换为计算机可以理解的语言。因此,为了实现商业规则,生成转换 112 以从一个或多个数据源,例如,输入数据集 100 接收输入记录(例如,记录 102)并且将指示乘客姓名 104 和组 118 的输出记录(例如,记录 114)产生到输出数据集 120 中。输入数据集和输出数据集还被称为数据流。

[0032] 为了对于非技术性的用户简化转换 112 的创建,典型地提供编辑器工具(未示出)以便以用户熟悉的格式输入被称作规则集或规则组的商业规则集。随后,所述规则组指导计算机系统生成转换 112,转换 112 更进一步指导计算机系统利用输入数据集 100 做什么以及产生什么到输出数据集 120 中。相应于单个转换的规则或规则集可以包括根据输入记录对于规则集的输出变量计算不同值的一个或多个规则例。当规则中的规则例被触发时,认为所述规则,更具体地,所述规则例被激发(fire)。例如,在规则中仅可以激发一个规则例。在某些例子中,在规则中可以激发多于一个的规则例。在某些例子中,当规则例被激发时,可以认为整个规则正在被激发。在一些实现方式中,例如,如果在输入数据集中的输入标量或矢量值满足规则例或规则中的一个或多个条件,则所述规则例或规则被触发或激发。规则集还可以包括其他规则集。其他规则集可以对于额外的或替换的输出变量产生值。例如,规则集可以直接地包含或间接地参考被称为“内含(included)”规则集的其他规则集。

[0033] 在图 2 中示出示例转换生成系统。生成器 150 从编辑器 154 接收规则集 152 作为输入并且生成转换 156。取决于系统结构以及转换和商业规则的目的,可以将生成的转换 156 提供给基于图的计算系统 158 作为用于图的组件或作为整个图本身。基于图的计算系统 158 可以提供允许程序员通过使用组件作为建立块(building block)来建立基于图的应用的计算环境。经常通过有向图来表示基于图的应用,其中图中的顶点表示组件(数据存储组件或可执行的计算组件二者之一),并且图中的有向链接(link)或“边”表示在组件之间的数据的流动。数据流图(也简称为“图”)是模块化实体。每个图可以由一个或多个其它图组成,并且具体的图可以是更大的图中的组件。

[0034] 例如,生成器 150 可以是被配置为使用标准工具以接收规则集 152 并且输出转换 156 的编译器、客户定制的程序或者基于图的计算。为那些本领域技术人员已知的用于产生以及随后更新转换 156 的任何技术都可用于生成转换 156。例如,在 2007 年 4 月 10 日提交的题目为“Editing and Compiling Business Rules”的美国专利申请 No. 11/733,434 中描述了用于产生转换的技术,通过引用将其全面包含于此。

[0035] 在某些例子中,转换 156 相应于输入记录 102 生成用于输出变量的仅一个值。在这种方案中,规则集最多仅可以激发一次。因此,使用转换 156 可能不能容易地解决一些问题,例如,数据质量问题。在某些例子中,输出数据集 120 中的输出变量可以包括“一次写入(Write-Once)输出”。通常,“一次写入输出”是对于给定输入记录典型地被写入一次,并

且对于给定输入记录仅存储一个值的输出变量。产生这种变量的规则集被称作“一次激发 (Fire-Once)”规则。

[0036] 在某些例子中,“多激发”规则可以产生“累积 (accumulator)”输出变量,例如,对于给定输入记录能够接收一系列的值,而不是仅一个值。对于输入记录,“多激发”规则将激发被触发的规则集中的每个规则例,而不仅仅是,例如,被触发的第一个规则例。

[0037] 在某些例子中,规则集可以以如图 3 所示的表格 (或“电子表格 (spreadsheet)”)格式被录入,其中在单元格中行和列交叉。表格 200 中的触发器列 202、204、206、208 相应于用于可得到的输入数据值的标准,并且行 210a-210h 相应于规则例,即,与可得到的输入数据值相关的标准的集合。处于触发器列和可应用的规则例行 210n 的交叉点的单元格包含用于触发列和规则例的标准。如果对于在其中规则例具有标准的各个触发器列来说,图 1 中的记录 102 的数据值满足触发标准,那么规则例 210n 就应用于给定的记录。如果应用了规则例 210n,则基于一个或多个输出列 212 生成输出。如上所述,通常,其所有的输入关系都得以满足的规则例可以称为“被触发 (triggered)”,并且规则集被称作“被激发 (fired)”。每个输出列 212 相应于一个潜在的输出变量,处于列 212 和可应用的行 210n 的交叉点的相应单元格中的值确定用于那个变量的输出——如果有的话。在某些例子中,单元格可以包含分配给变量的值或者它可以包含被评估以生成输出值的表达式,如下面将讨论的那样。虽然在图 3 中仅示出一个输出列,但是在某些例子中可以存在多于一个的输出列。

[0038] 可以存在若干不同类型的触发器列,包括相应于变量的列、包含表达式但是对其计算一次然后将其当作变量对待的列、以及仅包含表达式的列。仅包含表达式的列在某些方面比那些与变量相应的列或者当作变量对待的列简单。例如,这种触发器列可以包含如下类型的单元格值中的一个用于定义触发器列标准:

[0039] ●表达式。如果表达式评估为非零 (non-zero) 或非空 (non-NULL) 的值则该条件将被认为真。

[0040] ●关键字“任意 (any)”,或空串。该条件总为真。触发器列中的每个空单元格等于一个明确地包含关键字“任意”的单元格。

[0041] ●关键字“其它 (else)”。如果在该单元格的左边的全部单元格都相同的行中,在该单元格上面的包含“其它”的单元格都不为真,则该条件为真。

[0042] ●关键字“相同 (same)”。如果该单元格上面的单元格为真则该条件为真。

[0043] 相应于变量的列 (列变量) 可以具有两种类型的单元格。单元格的一种类型是表达式单元格。那些单元格完全像在仅包含表达式的列中的单元格一样动作 (behave)。然而,关键字“本身 (this)”可被用于表达式中以指向列变量。单元格的另一种类型是比较值。用于比较值的示例语法如下:

[0044] `comparison_value::= compound_value(“or” compound_value)*`

[0045] `compound_value::= simple_value(“and” simple_value)*`

[0046] `simple_value::= [“not”](value_expression|simple function|membership_expr)`

[0047] `value_expression::= [operator]value_element`

[0048] `operator::= “>” | “<” | “>=” | “<=” | “!= ” | “=” | “equals”`

[0049] `value_element::= constant|constant|variable|“(“expression“)”`

[0050] simple\_function:: = “is\_null” | “is\_blank” | “is\_valid” | “is\_defined” | “is\_bzero”

[0051] membership\_expr:: = “in” “[ “value\_element(( “,” | “to” | “or”) value\_element)\* “]”

[0052] 其中 “\*” 意味着一项 (term) 被重复零次或多次。

[0053] 可以使用任何适合的程序语言或语法。示例包括 C、Java、DML 或 Prolog。根据运算符、函数或从属关系 (membership) 表达式将列变量相对于比较值进行比较。在图 3 的示例中, 开头两个列 202 和 204 包含利用 “>=” 运算符的比较值。因此, 如果用于该列的值大于或等于相应的数字则标准被满足。如果没有运算符, 如在 “座位等级” 列中的那样, 则假定 “相等 (equal)”。常量可以是在基础系统中使用的任何编程语言或语法中的任何合法的常量。表达式是在使用的语言中的任何合法的表达式, 其返回将要相对于列变量进行比较的相容的数据类型。在某些例子中, 比较值内部的表达式被装入括号中以避免歧义。

[0054] 在图 3 的示例中, 第一行 210a 仅在一个列 202 中具有标准, 指示如果旅客的经常飞行英里的总数大于 1,000,000, 那么不管任何其它列可能具有什么值都应用该规则例。在该情况下, 针对该用户的 “乘机组 (Boarding Group)” 输出变量被设置为组 1。同样地, 第二规则例 210b 指示任何在头等舱的乘机者都在组 1 中。在一些例子中, 规则被顺次评估, 因此一个超过 1,000,000 英里且持头等舱票的旅客将在组 1 中, 但是将仅触发第一规则例 210a。

[0055] 规则例 210a-210h (图 3) 也可以表示为独立的简单规则, 各自在它们自己的表格中, 如图 4 所示。规则 220a-220d 分别相应于图 3 的行 210a-210d, 而规则 220e 具有相应于合起来的行 210e-210h 的四个规则例。用户可以分别地创建这些独立规则, 而非生成图 3 中示出的整个表格。每个规则例 (至少隐含地) 包含用于各个触发器列的值以及用于各个输出列的值 (该值可以为空, 即, 有效设置为 “任意”)。当多个规则生成相同的输出时, 将这些规则排序并且对它们按次序进行考虑, 直到针对输入触发一个规则中的规则例并且生成输出为止。如果没有触发规则中的规则例, 那么就处理生成相同输出的下一个规则。如果在任一规则中都没有例对于某一输出触发, 那么就使用默认值。

[0056] 在一些例子中, 编辑器的用户接口可以用于用图识别包含表达式的单元格。因此, 用户可以理解两个表达式之间的差异, 其中一个表达式将独立地被评估为真或假, 另一个表达式返回与列变量相比较的值。当用户打字时, 他可以例如通过在开头处打星号, 来指示特定单元格将是一个表达式单元格。

[0057] 对于相应于输出变量的列, 单元格可以包含下列中的一个:

[0058] ● 值。该值将分配给输出变量。

[0059] ● 表达式。该表达式的值将分配给输出变量。如果表达式评估为空 (NULL), 那么该字段得到空值, 除非输出字段不可为空。在输出字段不可为空但得到空值的情况下, 生成错误。

[0060] ● 关键字 “空 (null)”。如果输出字段可为空, 那么该字段将分配为空。否则, 生成错误。

[0061] ● 空字符串。如果输出字段具有默认值, 则分配该默认值。否则, 单元格被当作好像它包含关键字 “空 (null)” 一样对待。

[0062] ●关键字“相同 (same)”。输出字段被赋以在上面的单元格中所计算的相同的值。

[0063] 除表达式之外,可以允许用户向规则中的任何单元格附加注解,可以响应于用户交互(例如,单击或“悬停(hovering)”光标)来显示所述注释。

[0064] 在一些实现方式中,规则集,例如,在下面的表格 1 中示出的规则集可以包括对于单个输入记录生成多个输出记录的多个规则例。

[0065]

| 触发器 :汽车选择          | 触发器 :预算      | 输出 :配备等级              |
|--------------------|--------------|-----------------------|
| Honda S2000        | $\geq 37000$ | S2000CR               |
| Honda S2000        | 其它           | S2000                 |
| Honda Accord Coupe | $\geq 29000$ | Accord Coupe EX-L V-6 |
| Honda Accord Coupe | $\geq 26000$ | Accord Coupe EX-L     |
| Honda Civic Sedan  | $\geq 24000$ | Accord Coupe EX       |
| Honda Element      | 任意           | Accord Coupe          |
| ...                | ...          | ...                   |

[0066] 表格 1

[0067] 上面的规则集鉴于家庭预算考虑家庭的汽车选择,并且输出对于汽车的配备等级。在这种规则集(称为“标准化规则集”)的某些例子中,输出值的至少一个被确定为关键输出值,例如,“S2000CR”。当计算关键输出值“S2000CR”的规则被评估时,关于输入数据记录被触发以生成输出值“S2000CR”的规则例(汽车选择: Honda S2000 和预算:  $\geq 37000$ )被标注。随后在先前触发的规则例(汽车选项: Honda S2000 和预算:  $\geq 37000$ )被禁用的条件下该规则集被再次评估,以了解任一其他规则例是否触发并产生输出值。重复如上所述的处理直到没有另外的规则例被触发。每个输出值被存储为分开的输出记录。在某些例子中,规则例被分组,以使得如果一个规则例触发,则在被触发的规则例的组中的其它规则例也对于相同的输入记录关于接下来的重复被禁用。

[0068] 在某些例子中,相应于标准化规则集的转换可以使用两个阶段的处理。首先,读取输入记录并且通过,例如,调用“长度(length)”函数来计算计数(count)。计数相应于将生成的输出记录的数量。随后,对于每个输出记录调用另一函数,即,“标准化(normalize)”函数。标准化函数接收输入记录的副本和来自通过长度函数产生的计数的当前索引,并且产生输出值到不同的输出记录中。例如,如果输入记录具有四个(4)的家庭大小以及\$20,000的预算,则转换生成三个输出记录,用于三种建议的汽车(Accord Sedan、Civic 和 Element)中的每一种。

[0069] 在一些实现方式中,该转换对于汽车选择使用“长度”函数计算全部可能的值,以使得输出记录的数量已知。一旦该转换计算了全部可能的汽车输出值,该转换随后可以以与输出记录一样多的次数调用“标准化”函数,以向每个输出记录分配值。

[0070] 在一些实现方式中,不使用如上所述的两阶段程序,而是该转换可以通过若干次直接地调用“标准化”函数对于汽车选择计算全部可能的值直到再没有要计算的值。

[0071] 图 5 是用于对于输出变量 508 生成多个值 504 的示例规则集 500。用户可能对知道为什么特定的车辆被认为是无效 (invalid) 的所有的理由感兴趣,而不是仅知道第一个理由。在某些例子中,如图 6 所示,第一步是用户使用编辑器的用户接口中的输出表格片段 600 指定规则集 500 来产生多个输出值 504。

[0072] 因而,用户指示输出变量 508 “名称有效性消息 (Name Validation Message)”是用于接收一系列的值 504 的累积变量。相应于输出变量 508 的输出类型 604 改变为指示“累积器 (accumulator)”608。

[0073] 在某些例子中,相应于输出变量 508 的标量值可以是“累积的 (accumulated)”以供“记分卡 (score-card)”风格的规则集合使用。记分卡风格的规则集指的是用户指示正分数或负分数以被包括到规则值中的一种商业规则。因此,不是将相应于输出变量 508 的值存储为输出矢量,而是将相应于 508 的被累积的值的总和存储为标量值。

[0074] 在某些例子中,对于输出数据集中的每个记录,累积器输出变量 508 映射到可变长度矢量或阵列。因而,如果输出变量 508 被当作阵列,则用户可以为输出变量 508 指定大小。用户可以通过改变最大计数 (Max Count) 612 参数来指定输出变量 508 的长度。因此,字段 614 指示输出变量 508 被当作用于接收特定数量 (例如,20) 的值的阵列。在某些例子中,在用户没有指定尺寸的情况下,作为默认,输出变量 508 可以接收无限数量的值。因而,最大计数 612 参数指示,例如,“无限的 (unlimited)”。在某些例子中,为了帮助区分累积器类型输出变量和一次写入类型输出变量,编辑器可以禁止用户编辑用于一次写入变量的最大计数 612 参数。在某些例子中,如果用户从累积器输出变量切换到一次写入输出变量,则编辑器可以清空最大计数 612 参数。

[0075] 图 7 是示出多激发规则集,例如,“验证人员 (Validate Person)”的示例规则表格片段 700。通过注解“一次写入输出 (Write-Once Outputs)”或“累积器输出 (Accumulator Outputs)”将累积器输出 708 与一次写入输出 712 可视地区别开。此外,可能有多个其他注解。例如,可以在规则表格片段 700 的顶端 704 处指示规则集的类型,即,“多激发规则 (Fire-Many Rule)” (产生累积器输出的规则),或“一次激发规则 (Fire-Once Rule)” (产生标量输出的规则),或者在一边的垂直注解 712 指示“一次激发”或“多激发”。在某些例子中,不同的图标可以用于一次激发和多激发规则。在某些例子中,可以将被激发的所有的规则例高亮度显示以用于由用户检查。

[0076] 图 8 是示出累积器输出变量 801 “验证消息 (Validation Message)”的内容的示例结果表格片段 800。如所示,输出变量 801 可以对于每个记录假定第一系列的值 813,并且第一系列的值 813 中的至少一个值 (例如,相应于“TANGELA SCHEPP”的值) 可以假定显示为逗号分开的值的汇集的第二系列的值 816。在某些例子中,用户可以将鼠标指针“悬停”在累积器输出值之上以显现 (uncover) 示出累积的值的列表的工具提示 (tool tip)。在某些例子中,当执行包括,例如,基准 (benchmark) 数据的测试时,如果在基准数据中的矢量以任何方式不同于在输出中的矢量,则输出可以被标记为是不同的。例如,不同可以包括:基准矢量具有与输出矢量不同数量的项、基准矢量具有与输出矢量不同次序的项、以及每个矢量内的单独的项是不同的。

[0077] 如下所述,在操作中,累积器输出变量用于接收由多激发规则集产生的多个输出值。例如,考虑在表格 2 中示出的如下规则集:

[0078]

| 触发器:预算       | 触发器:家庭大小 | 输出:汽车选择            |
|--------------|----------|--------------------|
| $\geq 35000$ | $\leq 2$ | Honda S2000        |
| $\geq 22000$ | $\leq 2$ | Honda Accord Coupe |
| $\geq 20000$ | $\leq 4$ | Honda Accord Sedan |
| $\geq 15000$ | $\leq 4$ | Honda Civic Sedan  |
| $\geq 20000$ | $\leq 6$ | Honda Element      |
| $\geq 28000$ | $\leq 7$ | Honda Odyssey      |
| $\geq 50000$ | $\leq 4$ | Acura RL           |

[0079] 表格 2

[0080] 上面的规则集考虑,例如,家庭大小为 4 并且预算为 \$20,000,而建议三种车 (Accord Sedan、Civic 和 Element)。因此,在这种情况下,输出数据集中的输出变量“汽车选择”被认为能够接收多个值。该规则集中的每个规则例都被评估并且任何时候规则例触发时,来自该上面的规则集的值被添加到累积器输出变量。

[0081] 上面的规则集中的触发器可以是包括输入值、查找及其他输出值的任一标量变量 (非矢量)。在某些例子中,输出变量可以计算另一输出变量。在某些例子中,仅非矢量输出可被用作触发器。在某些例子中,有可能通过使用聚集函数 (aggregation function) 来间接地使用一个累积器输出变量以计算另一累积器输出变量。例如,考虑在表格 3 中示出的如下规则集:

[0082]

| 触发器    | 输出:家庭成员 |
|--------|---------|
| 在世     | 自己      |
| 已婚且未分居 | 配偶      |
| 有婴儿    | 婴儿      |
| 有未成年女儿 | 女儿      |
| 有未成年儿子 | 儿子      |

[0083] 表格 3

[0084] 上面的规则集计算称作“家庭成员”的累积器输出变量。现在,考虑在表格 4 中示

出的如下规则集：

[0085]

|          |
|----------|
| 输出：家庭大小  |
| 计数（家属成员） |

[0086] 表格 4

[0087] 表格 4 中的规则集使用聚集函数计算称作“家庭大小”的标量（非矢量）。因此，首先，计算包括全部我们的家庭成员的列表的输出矢量。随后计数函数计算列表中的人的数量。该计数随后用作输入以计算汽车的列表。

[0088] 图 9 示出使用标量和矢量以对于使用累积器输出变量的其他标量和矢量计算值的示例实现。如所示，S1，S2 和 S3 表示标量变量。V1 和 V2 表示矢量变量。S1 用来计算 S2；随后 S2 用来计算 V1 的四个不同的值。随后 V1 的全部四个值用来计算 S3（例如，通过使用聚集函数）。最后，S3 用来计算 V2 的三个值。

[0089] 在一些实现方式中，当用户试图执行如下示例动作中的任何一个动作时编辑器可以产生有效性错误：当在任一数据集中的字段的类型是任何除可变长度矢量以外的类型时标记输出为累积器；当在任一数据集中的字段的类型是矢量时标记输出为“一次写入”；（在仅一次写入输出可以具有默认值的实现中）对于累积器提供默认值，使用累积器输出作为比较触发器列；在单个规则内混合累积器和一次写入输出；以及在累积器输出变量的最大计数参数中输入除无限或正数以外的值。

[0090] 在某些例子中，输入记录可以包括矢量。图 10A 是输入记录 950 的示例格式，输入记录 950 包括至少两个矢量记录，即，驾驶员记录矢量 952 和车辆记录矢量 954。图 10B 示出用于输入记录 950 的示例数据 956。

[0091] 聚集函数可以被包括在规则集中以将记录矢量 952、954 变换为标量。例如，规则集可以包括规范“最年轻的驾驶员的年龄”。在一些实现方式中，规范表示为“最小（驾驶员年纪）”，或可以使用诸如“do minimum(in0.drivers, ‘age’)”的数据操作语言（DML）。响应于规则集，例如，（从图 10B 中的 Pebbles 的记录）产生标量值 21。在某些例子中，在操作中，函数可以循环通过驾驶员记录矢量 952 中的全部记录以找到驾驶员年龄的最小值。

[0092] 考虑另一示例，规则集中的规范可以是“对于最年轻的男性驾驶员的点数（points）加一的数”。该规范可以表示为“最小（驾驶员年龄，驾驶员性别=男性，驾驶员点数+1）”。响应于此规则集，例如，（从 BamBam 的记录）产生标量值 14。在一些实现方式中，该标量值可以被分配给为标量的中间变量或输出变量。

[0093] 在某些例子中，可以对于在记录矢量中的每个元素写入规则。例如，考虑在表格 5 中示出的如下规则集：

[0094]

|             |              |         |
|-------------|--------------|---------|
| 车辆具有气囊（触发器） | 车辆具有安全带（触发器） | 值调整（输出） |
| 否           | 否            | 0       |

|   |   |     |
|---|---|-----|
| 否 | 是 | 100 |
| 是 | 否 | 150 |
| 是 | 是 | 300 |

[0095] 表格 5

[0096] 表格 5 的规则集中的规范是“对于每个汽车,对汽车的值计算调整,如果汽车有安全带则该调整为 100,如果汽车有气囊则该调整为 150,以及如果汽车有安全带和气囊两者则该调整为 300”。如所示,输出变量“值调整”是矢量变量。响应于上述规则,产生矢量,例如, [0,300,100]。在某些例子中,在操作中,多次运行规则集,对于车辆记录矢量 954 中的每个记录都运行一次。

[0097] 在某些例子中,规则集还可以参考标量值,或者在其他矢量具有相同的长度的时候参考其它矢量。例如,考虑在表格 6 中示出的如下规则集:

[0098]

| 车辆车龄 | 调整的值(输出)             |
|------|----------------------|
| > 2  | 车辆值 + 值调整 + 地理调整 -50 |
| 其它   | 车辆值 + 值调整 + 地理调整     |

[0099] 表格 6

[0100] 表格 6 的规则集中的规范是“对于每个汽车,计算调整的值,该调整的值是汽车值、它的值调整和地理 (geographic) 危险性的总和。如果该汽车是大于两年的旧车,则减去 50”。在此规则中,“调整的值”是矢量变量。因此,为了避免由于不相等的矢量长度导致的运行时错误,矢量变量“值调整”具有与车辆记录矢量 954 相同的长度。响应于此规则集,产生矢量,例如, [1030,1880,1330]。

[0101] 在某些例子中,当 XML 记录是复数 (complex) 时,单个输入记录可用于通过利用关键字关系关联逻辑记录来表示多个逻辑记录。例如,车辆记录矢量 954 中的每个车辆子记录可以包括外关键字,例如,“驾驶员”,以关联到驾驶员记录矢量 952 中的匹配关键字,例如,“姓名”。用这样的方式,记录矢量 952,954 可以实现为查找文件或内部参考表格。例如,与车辆记录矢量 954 相联系的内部参考表格可以如下:

[0102] Primary Driver Name(primary key)(主驾驶员姓名(主关键字))

[0103] Primary Driver Age(主驾驶员年龄)

[0104] Primary Driver Sex(主驾驶员性别)

[0105] Primary Driver Points(主驾驶员点数)

[0106] 因此,可以通过将记录矢量中的子记录当作内部参考表格中的记录来对于每个输入记录创建内部参考表格。在操作中,考虑在表格 7 中示出的如下规则集:

[0107]

|                  |
|------------------|
| 输出 :投保驾驶员的年龄     |
| 主驾驶员年龄 ( 投保驾驶员 ) |

[0108] 表格 7

[0109] 表格 7 的规则集中的规范是是“计算投保驾驶员 (Policy Driver) 的年龄,该年龄是通过使用投保驾驶员的值作为用于关联内部参考表格的关键字找到的投保驾驶员的年龄”。该规范在主驾驶员年龄列返回值,该值随后被分配给输出变量,“投保驾驶员”。“投保驾驶员”是标量值。在另一示例中,考虑下面的表格 8 中示出的规则集：

[0110]

|                |
|----------------|
| 购买时的年龄 ( 输出 )  |
| 主驾驶员的年龄 - 车辆车龄 |

[0111] 表格 8

[0112] 表格 8 的规则集中的规范是“计算购买时的年龄 (Age at Purchase),该值是车辆的年龄与车辆的主驾驶员的年龄之间的差”。为了说明,假定默认地将“车辆主驾驶员”分配给查找关键字。输出变量“购买时的年龄”是矢量变量。因此,响应于上述规则,产生 [3119, 27]。

[0113] 在某些例子中,可以在括号中明确地指定查找关键字“车辆主驾驶员”为如下的“主驾驶员的年龄 ( 车辆主驾驶员 )- 车辆车龄”。

[0114] 在某些例子中,可以在聚集函数中使用内部参考表格。例如,规范可以是“在所有车辆的主驾驶员的年龄上计算平均值”。此规范可以通过函数,例如,“平均 ( 主驾驶员的年龄 ( 车辆主驾驶员 ))”实现。响应于此函数,产生标量值,例如,29. 67。

[0115] 在一些实现方式中,用户可以使上述规则集中的计算步骤具体化。例如,在测试方式中,能够检查感兴趣的值对用户来说是有用的,例如,检查输入变量与输出变量 ( 两者都是标量和两者都是矢量变量 ) 的中间值。可以使用本领域已知的各种技术来实现该具体化。例如,可以实现具有用于输入记录矢量 952、954 中的每个元素的行的弹出表格以汇总指示什么项已经被滤出或计算的中间值。

[0116] 可以使用计算机上运行的软件实现如上所述的技术。例如,所述软件在一个或多个编程或可编程计算机系统 ( 可以是诸如分布式、客户机 / 服务器或网格之类的各种架构 ) 上运行的一个或多个计算机程序中形成过程,所述计算机系统的每一个都包括至少一个处理器、至少一个数据存储系统 ( 包括易失的和非易失性存储器和 / 或存储元件 )、至少一个输入设备或端口以及至少一个输出设备或端口。所述软件可以形成大型程序的一个或多个模块,例如,所述大型程序提供与计算图的设计和构造相关的其他业务。图的节点和元素可以实现为存储在计算机可读介质中的数据结构或符合存储在数据容器中的数据模型的其他组织化的数据。

[0117] 所述软件可以提供在诸如 CD-ROM 的存储介质上,该存储介质可以由通用或专用可编程计算机读取或者 ( 以传送信号编码 ) 通过网络的通信介质传递给软件运行的计算机。所有的函数可以在专用计算机上执行,或者使用诸如协处理器的专用的硬件。所述软

件可以以不同的计算机执行由软件指定的计算的不同的部分的分布模式实现。每个这种计算机程序优选地存储在或装入到通用或专用可编程计算机可读的存储介质或设备（例如，固态存储器或介质，或者磁性或光学介质）上，当由计算机系统读取存储介质或设备以执行此处描述的过程时所述计算机程序用于配置和操作计算机。本发明系统还可以被认为是实现为利用计算机程序配置的计算机可读存储介质，这样配置所述存储介质导致计算机系统以特定的和预先确定的方式操作以执行此处描述的函数。

[0118] 已经描述本发明的大量实施例。然而，清楚地是，可以做出各种修改而不脱离本发明的精神和范围。例如，如上所述的一些步骤可以是独立的次序，从而可以以不同于描述的次序执行。

[0119] 还应理解地是，上文的描述旨在示出而非限制本发明的范围，本发明的范围由所附的权利要求的范围限定。例如，可以以不同的次序执行上面描述的大量函数步骤而不实质上影响全过程。其他实施例在本申请的权利要求书的范围内。

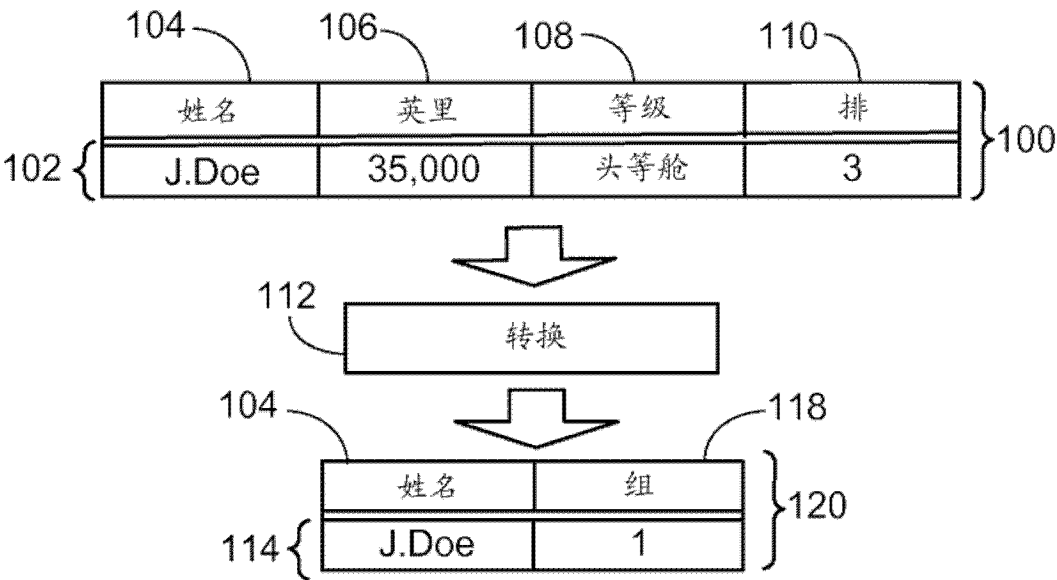


图 1

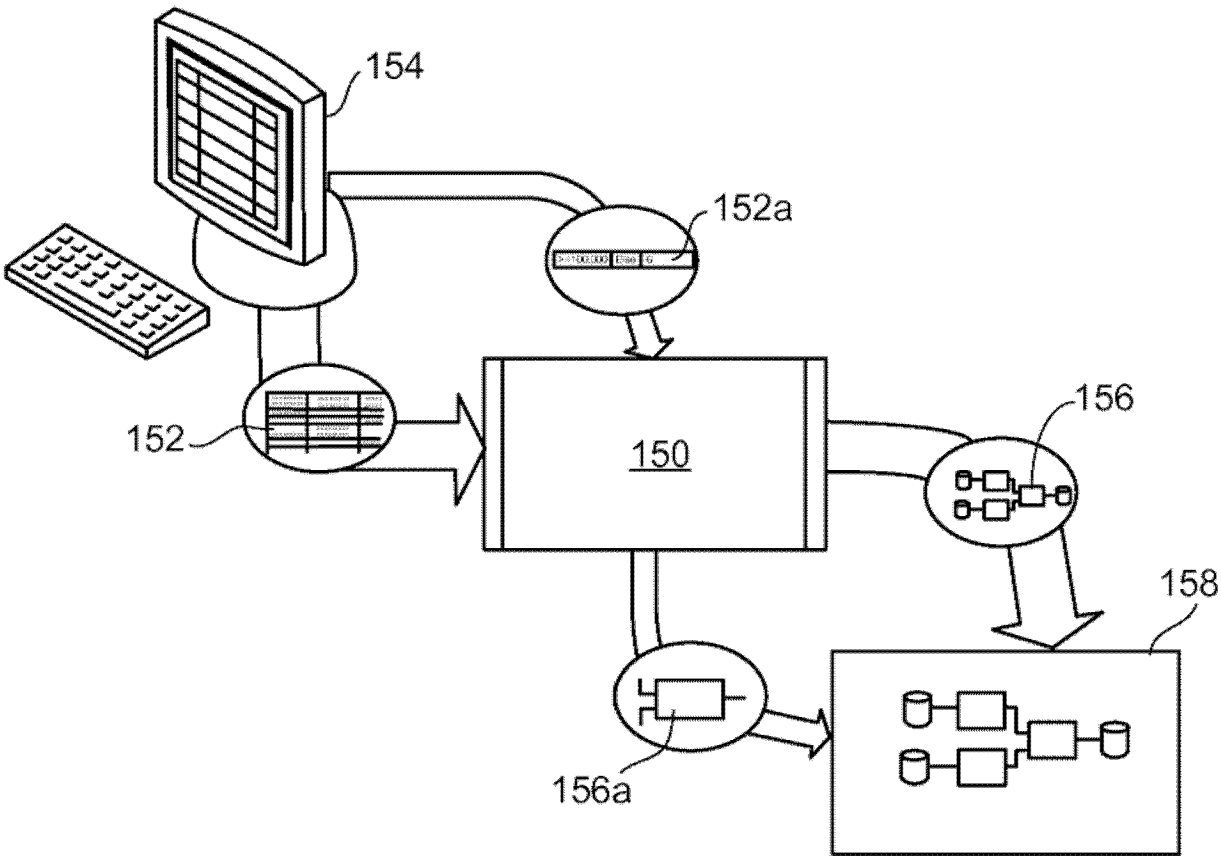


图 2

|      | 202           | 204              | 206  | 208  | 212 | 152 |
|------|---------------|------------------|------|------|-----|-----|
|      | 经常飞行<br>英里的总数 | 本年度经常飞行英里        | 座位等级 | 座位排  | 登机组 |     |
| 210a | >=1,000,000   |                  |      |      | 1   |     |
| 210b |               |                  | 头等舱  |      | 1   |     |
| 210c | >=100,000     | >= 上年度经常<br>飞行英里 |      |      | 2   |     |
| 210d |               |                  | 商务舱  |      | 2   |     |
| 210e |               |                  | 其它   | <=10 | 2   |     |
| 210f |               |                  | 相同   | <=40 | 3   |     |
| 210g |               |                  | 相同   | <=50 |     |     |
| 210h |               |                  | 相同   | 其它   | 5   |     |

200

图 3

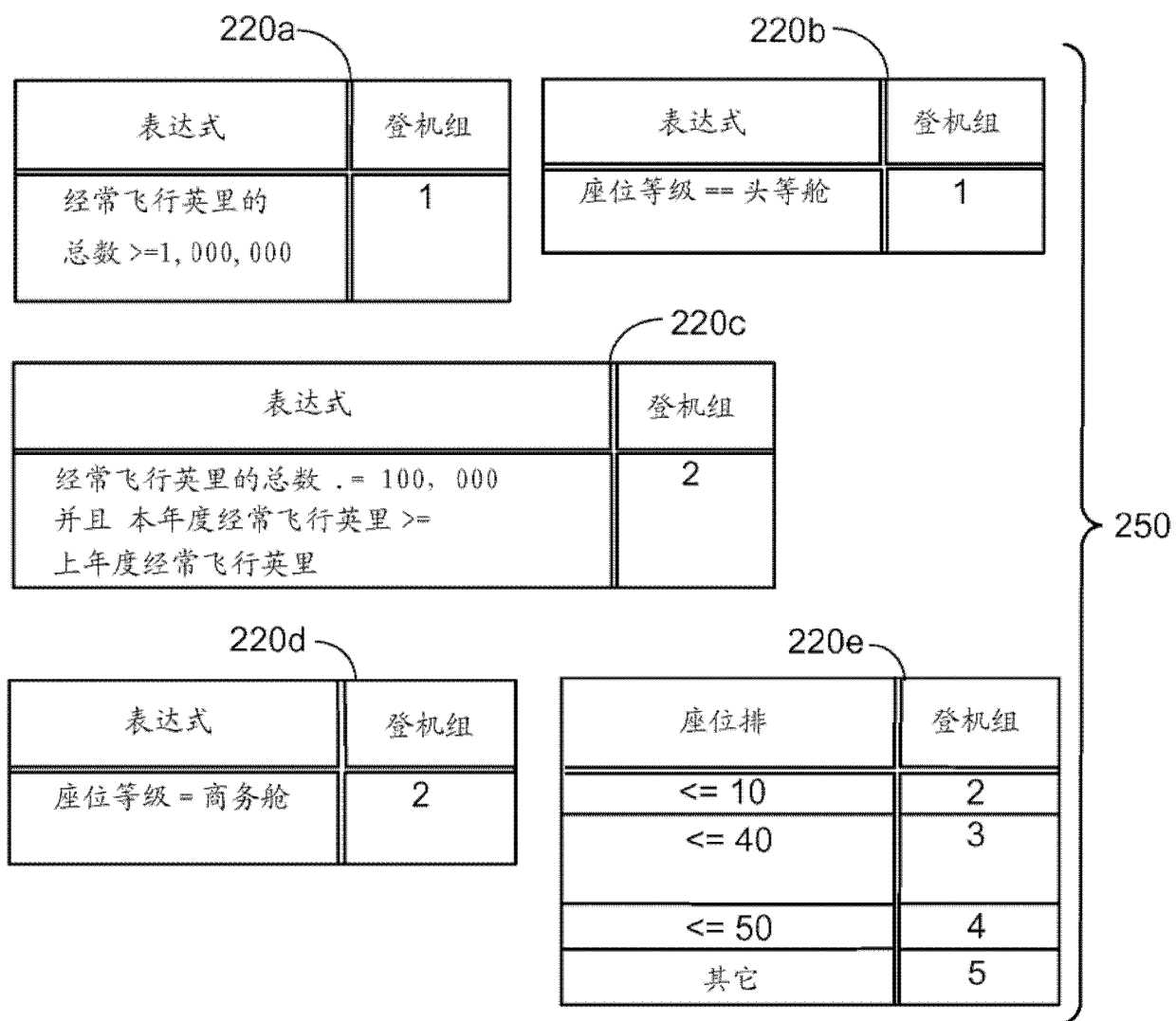


图 4

| 500                                |       |        |      |                  |      |
|------------------------------------|-------|--------|------|------------------|------|
| 508                                |       |        |      |                  |      |
| 输出                                 |       |        |      |                  |      |
| 触发器                                |       |        |      |                  |      |
|                                    | 年代    | 车辆值    | 质优车辆 | 如果为真             | 有效车辆 |
| 1                                  | >2004 | <15000 | 否    | *Seat Material = | 否    |
| 2                                  |       | ↓      | 任意   | *Four Wheel...   |      |
| 3                                  |       | 任意     | 是    | *Power Wind...   |      |
| 4                                  | 任意    |        | 任意   | *On Star = Yes   |      |
| 5                                  | <2000 | 任意     | 任意   | *Ipod Jack = Ye  |      |
| 6                                  | <1990 | 任意     | 任意   | *CD Changer = Y  |      |
| 7                                  | <2004 | 任意     | 任意   | *Blue Tooth E... |      |
| 8                                  | 其它    | 任意     | 是    | ↓                |      |
| 9                                  | <1995 | 任意     | 任意   | *Car Color in M  |      |
| 10                                 | <1900 | 任意     | 任意   | 任意               |      |
| 11                                 | 任意    | 任意     | 任意   | 任意               | 是    |
| *                                  |       |        |      |                  | 默认值  |
| 504                                |       |        |      |                  |      |
| 车辆验证消息                             |       |        |      |                  |      |
| "Vehicle value too low to have.."  |       |        |      |                  |      |
| "Vehicle value too low to have.."  |       |        |      |                  |      |
| "Recent premium vehicles al..."    |       |        |      |                  |      |
| "Vehicle value too low to have.."  |       |        |      |                  |      |
| "Vehicle too old to have had a..." |       |        |      |                  |      |
| "Vehicle too old to have had a..." |       |        |      |                  |      |
| "Vehicle to old to have had a..."  |       |        |      |                  |      |
| "Recent premium vehicles ha..."    |       |        |      |                  |      |
| "Vehicle too old to have had ...." |       |        |      |                  |      |
| "Vehicle too old to insure."       |       |        |      |                  |      |

图 5

| 600 |                                 | 604 |      |     |      |      | 612                        |    |
|-----|---------------------------------|-----|------|-----|------|------|----------------------------|----|
|     | 姓名                              | 类型  | 常数名称 | 默认值 | 输出类型 | 最大计数 | 在数据集中<br>的技术名称：<br>“数据集 1” | 描述 |
| 1   | <input type="checkbox"/> 姓名有效   | 串   | 0    | “是” | 一次写入 |      | name_valid                 |    |
| 2   | <input type="checkbox"/> 姓名验证消息 | 串   | 0    | 空   | 累加器  | 20   | name_valid                 |    |
|     | 新变量                             |     |      |     |      |      |                            |    |

图 6

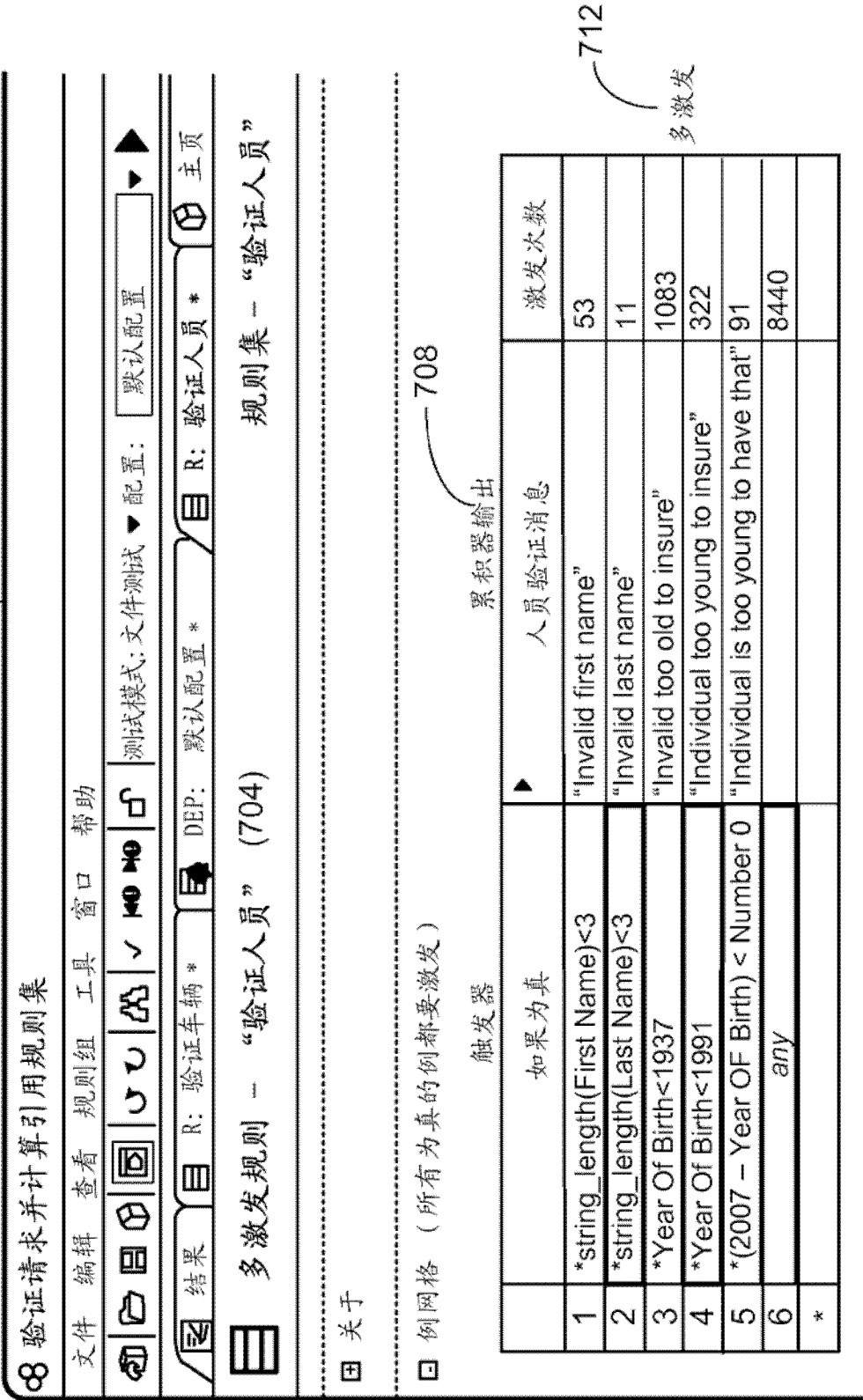


图 7

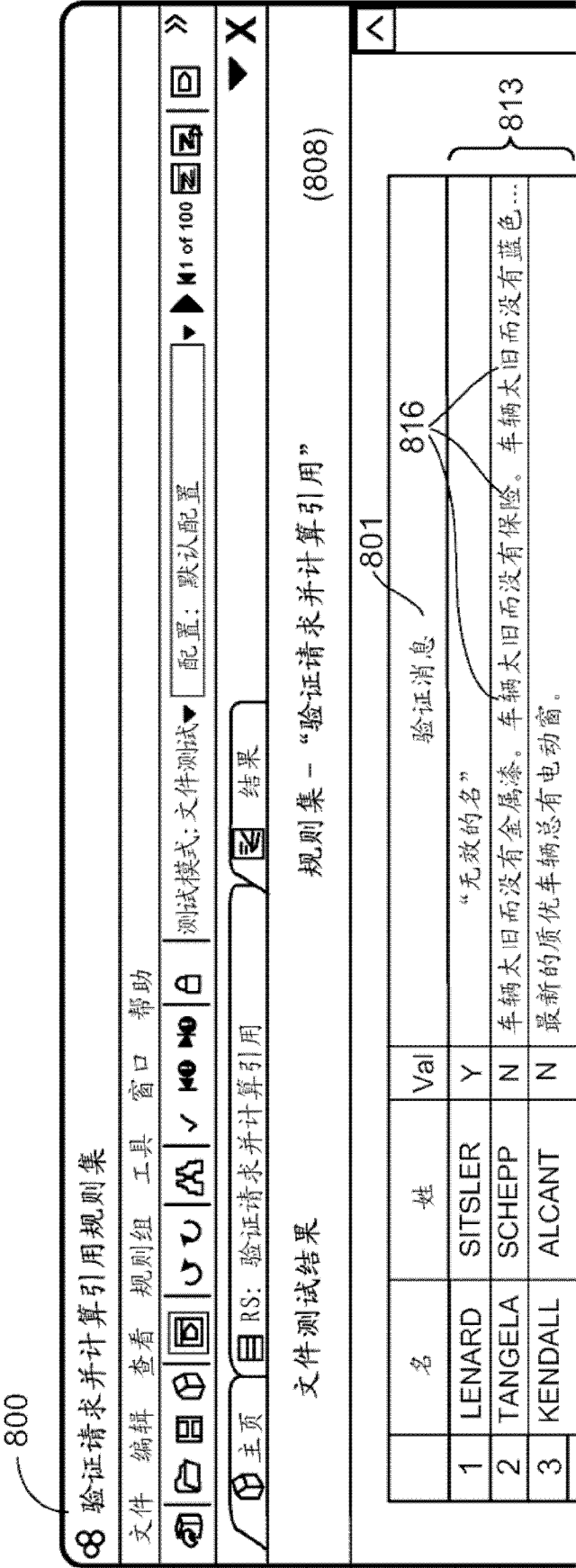


图 8

22

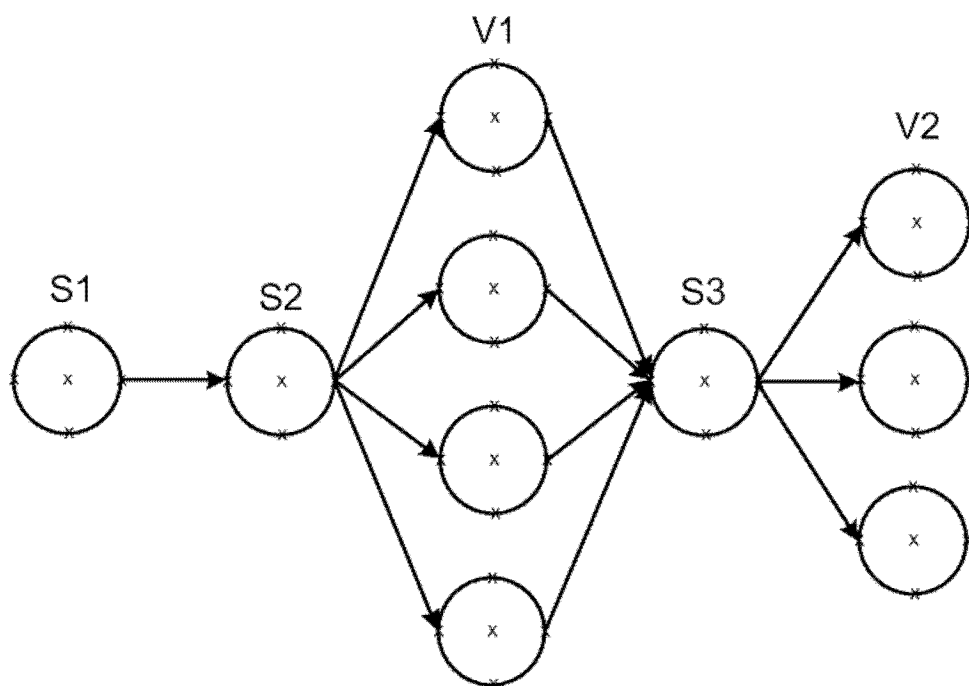


图 9

```

record
    decimal 1 (2) num_drivers; // 驾驶员数量
    record
        string (20) name; // 驾驶员姓名
        string (1) sex; // 驾驶员性别
        decimal (3) age; // 驾驶员年龄
        decimal (2) points; // 驾驶员点数
    end driver [num_drivers];
    decimal (2) num_vehicles; // 车辆数量
    record
        decimal (10) value; // 车辆值
        decimal (2) age; // 车辆车龄
        decimal (1) airbag; // 车辆具有气囊
        decimal (1) has_belt; // 车辆具有安全带
        decimal (20) driver; // 车辆主驾驶员
    end driver [num_vehicles];
    decimal (2) geo_risk; // 地理调整
end;
  
```

952 {

954 {

950

图 10A

956

```
<record>
<drivers>
  <driver name="Fred" sex="M" age="35" points="12" />
  <driver name="Wilma" sex="F" age="33" points="9" />
  <driver name="Pebbles" sex="F" age="21" points="14" />
  <driver name="BamBam" sex="M" age="22" points="13" />
</drivers>
<vehicles>
  <vehicle value="1000" age="4" air_bag="0" has_belt="0" drivers="Fred" />
  <vehicle value="1500" age="2" air_bag="1" has_belt="1" drivers="Pebbles" />
  <vehicle value="1200" age="6" air_bag="0" has_belt="1" drivers="Wilma" />
</vehicles>
<geo_risk value="80" />>
</record>
```

图 10B