



[12] 发明专利申请公开说明书

[21] 申请号 02803281.0

[43] 公开日 2004年3月3日

[11] 公开号 CN 1479902A

[22] 申请日 2002.9.30 [21] 申请号 02803281.0

[30] 优先权

[32] 2001.10.18 [33] KR [31] 2001-0064477

[86] 国际申请 PCT/KR02/01834 2002.9.30

[87] 国际公布 WO03/034287 英 2003.4.24

[85] 进入国家阶段日期 2003.6.23

[71] 申请人 翰迪斯福特公司

地址 韩国汉城

[72] 发明人 金荣镐 金昇 郭柄权 申容云

[74] 专利代理机构 中国国际贸易促进委员会专利商
标事务所

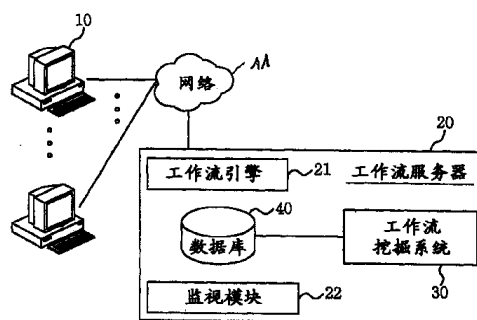
代理人 李强

权利要求书 2 页 说明书 12 页 附图 24 页

[54] 发明名称 workflow挖掘系统和方法

[57] 摘要

本发明披露了一种通过在工作流系统的运行过程中累积的工作流记录数据中使用特定的数据发掘技术，而可以评价、分析和确定过程或活动的先前执行结果的工作流系统，以及由此得到的方法。 workflow挖掘系统从 workflow服务器的数据库中提取必要的数据库，生成分析表，执行一个用于在所提取的数据的基础上去除不必要的属性或者将数字变量转化成符号变量的预处理过程，并且通过使用经过预处理的数据来分析决策树或关联性。 C4.5 算法被用于决策树分析， Apriori 算法被用于关联性分析。



1. 一种安装在具有一个 workflow 引擎和一个数据库的 workflow 服务器上的 workflow 挖掘系统，用于基于数据库的记录数据进行数据挖掘，包括：

一个用于从数据库中提取必要的数 据并生成分析表的数据提取和处理装置；

一个用于在所提取的数据的基础上除去不必要的属性或者将数字变量转换成符号变量的预处理装置；

一个用于分析决策树，通过使用经过预处理的数据来推断决策树类型的可能规则，或者在属性之间的支持和可靠性的基础上分析关联性的分析装置。

2. 如权利要求 1 所述的系统，还包括一个用来在经过预处理的数据的基础上，以文本或图形的形式为数据集提供统计值的数据统计模块。

3. 如权利要求 1 所述的系统，其中的分析表包括一个过程分析表和一个活动分析表。

4. 如权利要求 1 所述的系统，其中的预处理装置包括一个用于装载分析表的数据的分析表数据装载装置。

5. 如权利要求 1 所述的系统，其中用于分析决策树的算法是 C4.5 算法。

6. 如权利要求 1 所述的系统，其中用于分析关联性的算法是 Apriori 算法。

7. 一种在 workflow 服务器的数据库中的记录数据的基础上进行数据挖掘的 workflow 挖掘方法，包括：

(a) 一个用于从数据库中提取必要的数 据并生成分析表的步骤；

(b) 一个用于在所提取的数据的基础上除去不必要的属性或者将数字变量转换成符号变量的预处理步骤；以及

(c) 一个用于分析决策树，通过使用经过预处理的数据来推断决

策树类型的可能规则，或者在属性之间的支持和可靠性的基础上分析关联性的分析步骤。

8. 如权利要求 7 所述的方法，还包括一个在步骤(b)和(c)之间在经过预处理的数据的基础上，以文本或图形的形式为数据集提供统计值的步骤。

9. 如权利要求 7 所述的方法，其中的分析表包括一个过程分析表和一个活动分析表。

10. 如权利要求 7 所述的方法，其中的预处理步骤包括一个用于装载分析表的数据的分析表数据装载步骤。

11. 如权利要求 7 所述的方法，其中用于分析决策树的算法是 C4.5 算法。

12. 如权利要求 7 所述的方法，其中用于分析关联性的算法是 Apriori 算法。

13. 如权利要求 7 所述的方法，其中当关联性分析在步骤(c)中执行时，用于将数字变量转化为符号变量的过程总是在步骤(b)中执行。

workflow挖掘系统和方法

技术领域

本发明涉及 workflow挖掘，特别是涉及一种 workflow挖掘系统和方法，本系统和方法通过对在 workflow系统的运行过程中累积的 workflow记录数据以及由连接到 workflow系统的应用系统所累积的数据应用一种数据挖掘技术，能够评价、分析和确定过程或活动的先前执行结果。

背景技术

通常， workflow被定义为根据一系列的程序或规则将文件、信息或活动从一个参与者处传输到另一个参与者处的整个或部分商务过程。 workflow系统是一个用于集中管理 workflow的自动系统。

典型的商务过程包括保险公司的保险合同业务工作、递送公司的递送业务工作以及银行的贷款业务工作。 workflow系统可以解释和具体化包括与一个组织的日常工作相关的资源之间的关联性和流程的商务过程，因此人们可以通过使用商务过程来工作，它被称为过程执行。

现在参考图 1 对保险合同业务工作进行详细的解释。此商务过程具有连续的步骤，即接受客户订单、准备合同文件、批准准备的文件。在保险合同业务工作中，订单接受、文件准备、文件批准工作是 workflow中的各个单位活动。它是组成过程的最小逻辑单元，包括人力、资源和信息。

当接受了一个保险合同订单时，根据过程定义所产生的过程就是一个实例。因此，在一个过程定义中存在大量的实例。当一个实例被执行并结束时，与此过程实例相关的信息被存储在一个档案文件中。

workflow系统通常提供了对一个过程及在此过程中工作的一个

人，进行搜索、分析或评价的功能，包括一个开始和一个结束时间、一个负责人、一个耽搁工作的人、一个延误的工作和一个完成整个过程所需的时间。 workflow 系统的这些功能用来实施经济的技术，例如基于活动的管理或平衡计分卡。

但是，传统的工作流系统只能提供用于评价活动或过程的非常基本的功能。结果，出现了对新分析方法的强烈要求，此分析方法通过对在工作流执行过程中累积的大量数据进行有效的分析来度量和评价活动过程的结果。

发明内容

因此，本发明的一个目标是提供一种设计用来在工作流系统中使用一种特殊数据挖掘技术的工作流挖掘系统和方法。

为了实现上述发明目标，这里提供了一种安装在具有一个 workflow 引擎和一个数据库的 workflow 服务器上的 workflow 挖掘系统，用于根据数据库的记录数据进行数据挖掘，包括：一个用于从数据库中提取必要的数 据并生成分析表的数据提取和处理装置；一个用于在所提取的数据的基础上去除不必要的属性或者将数字变量转换成符号变量的预处理装置；以及一个用于分析决策树，通过使用经过预处理的数据来推断决策树类型的可能规则，或者在属性之间的支持和可靠性的基础上分析关联性的分析装置。

另外，这里提供了一种在工作流服务器的数据库中的记录数据的基础上进行数据挖掘的工作流挖掘方法，包括：(a)一个用于从数据库中提取必要数据并生成分析表的步骤；(b)一个用于在所提取的数据的基础上去除不必要的属性或者将数字变量转换成符号变量的预处理步骤；以及(c)一个用于分析决策树，通过使用经过预处理的数据来推断决策树类型的可能规则，或者在属性之间的支持和可靠性的基础上分析关联性的分析步骤。

附图说明

参考附图可以更好地理解本发明，这些附图只是用于示例，而并不是对本发明的限制，其中：

图 1 是一个描述了一个用于保险合同业务工作的商务过程的视图；

图 2 是一个描述了一个根据本发明的工作流系统的应用情况的视图；

图 3 是一个描述了 workflow 挖掘系统的详细结构视图；

图 4A 至 4D 是描述了数据提取和处理模块的一个处理过程的视图；

图 5A 至 5D 是描述了预处理模块的一个处理过程的视图；

图 6A 至 6C 是描述了数据统计模块的一个处理过程的视图；

图 7A 至 7E 是描述了决策树分析模块的一个处理过程的视图；

图 8A 至 8D 是描述了关联性分析模块的一个处理过程的视图；

以及

图 9 是一个示出了一个根据本发明的工作流挖掘方法的流程图。

具体实施方式

现在将参考附图详细描述本发明的工作流挖掘系统及相应的方法的一个优选实施例。

图 2 是一个描述了本发明的工作流挖掘系统 30 的一个应用情况的视图。参考图 2，workflow 挖掘系统 30 安装在一个 workflow 服务器 20 上，并且 workflow 服务器 20 与至少一个客户计算机 30 相连。这些系统可以通过一个内部的局域网（LAN）、广域网（WAN）、或互联网而互相访问。根据本发明，“互联网”的意思是根据一种标准协议，如 TCP/IP、HTTP，和一种无线互联网协议（例如 WAP）可以互相访问的网络集合。它还包括以后的变化，包括对现有标准协议的修改和补充。

workflow服务器 20 包括一个 workflow引擎 21、一个监视模块 22 和 workflow挖掘系统 30。

workflow引擎 21 根据过程定义控制活动，并且接收用于从客户端增加、校正、删除和搜索某个文档文件的控制命令，或者根据一个自操作来操作。 workflow引擎 21 包括一个组织管理模块、一个文件夹管理模块、一个流程控制模块、一个 DB 存储模块和一个文件管理模块。

组织管理模块从事 ID 管理，以及访问权授予和去除（登录控制），以管理加入 workflow系统的人。

文件夹管理模块用于系统化地管理关于过程定义的项目。即，它使得用户可以知道在哪个文件夹中存在哪个过程定义。例如，车保险过程定义、人寿保险过程定义和火灾保险过程定义可以被存储在一个保险文件夹中。

流程控制模块用于根据来自客户计算机 10 的命令或自操作将控制命令传送给 DB 存储模块或文件管理模块。

DB 存储模块在数据库 40 中存储了在客户计算机 10 中定义的一个文档文件、过程定义、文档文件之间的版本关联性、以及文档文件的校正日期和校正者。

监视模块 22 包含根据由流程控制模块生成的各种过程信息来观察处理过程的流程的功能。这里，各种过程信息存储在数据库 40 中。

数据库 40 存储了在工作流系统的操作过程中所生成的各种记录数据，如通过操作 workflow引擎 21 而生成的过程定义、档案文件、实例、用户、文档文件、文档历史、创建者和校正者。

workflow挖掘系统 30 在管理员的请求下，根据数据库 40 中的记录数据来评价、分析和确定过程和活动的先前执行结果，因此该数据可以用作商务过程重调整（BPR, business process reengineering）的客观数据。

图 3 是一个描述了 workflow挖掘系统 30 的详细结构的视图。如图

3 所示， workflow 挖掘系统 30 包括一个数据提取和处理模块 32、一个预处理模块 34、一个数据统计模块 36 和一个分析模块 38。

数据提取和处理模块 32 从数据库 40 中提取必要的数据库，并生成一个分析表。可以对过程和活动进行分别分析。通常，首先分析过程，然后再详细的分析活动，但并不仅限于此。数据提取和处理模块 32 具有一个过程分析表的生成模块和一个活动分析表的生成模块，以提取数据并生成分析表。预处理模块 34 根据由分析表提取的数据搜索数据的特性，去处不必要的属性，如果必要则划分实例，并且通过对区域 (section) 进行划分将数字变量转化为符号变量。为此，预处理模块 34 包括一个分析表数据装载模块、一个属性选择和去除模块和一个过滤模块，过滤模块包括一个离散化过滤器、一个记录过滤器和一个遗漏值过滤器。

数据统计模块 36 以文本或图形的方式为经过预处理的数据集提供基本的统计值，以掌握变量之间的基本关联性。为此，数据统计模块 36 包括一个预处理数据装载模块和一个显示模块。

分析模块 38 通过使用一些分析算法来对经过预处理的数据进行分析。分析模块 38 对决策树和关联性进行分析，因此它包括一个决策树分析模块和一个关联性分析模块。

图 4A 至 4D 是描述了数据提取和处理模块 32 的一个处理过程的视图；

图 4A 示出了一个初始屏幕，其中数据提取和处理模块 32 生成分析表。如图 4A 所示当按下 “ WM_PROCESS TABLE GENERATION ” 按钮而生成过程分析表时，如图 4B 所示过程分析表生成模块被执行，用于连接数据库 40 的输入窗口被打开，在其中输入用户 ID、密码以及分析对象的过程定义 ID。分析表生成模块从数据库 40 中提取具有所输入的过程定义 ID 的过程实例的一般信息，并完成表的生成。图 4C 示出了所生成的分析表的一个例子。如图 4C 所示，过程分析表的属性包括一个服务器 ID (SVRID)、一个过程实例的标识符 (PROCID)、一个表示紧急的标记 (URGENT)、一

个表示过程项中的密码验证状态的标记 (PASSWDF)、一个注释号码 (CMNTCNT)、一个附件号码 (ATTACHCNT)、一个过程实例的解释名称 (NAME)、一个 a.m./p.m. 形式的创建时间 (CREATIONTIME1)、一个在一周的某一天的创建时间 (CREATIONTIME2)、一个在时间间隙中的创建时间 (CREATIONTIME3)、一个过程实例的创建者 ID (CREATOR) 和一个过程实例的创建者的名称 (CREATOR NAME)。因为以前准备在数据库 40 中的记录数据被用作属性的数据, 所以过程分析表的属性可以是数据库 40 的属性的一部分或全部。

另一方面, 当数据提取和处理模块 32 试图生成活动分析表时, 在图 4A 的初始屏幕中按下 “WM_ACTIVITY TABLE GENERATION” 按钮, 如图 4D 所示活动分析表生成模块被执行, 用于连接数据库 40 的输入窗口被打开, 在其中输入用户 ID、密码、以及分析对象活动的过程定义 ID 和活动顺序 ID。活动表生成模块从数据库 40 中提取具有所输入的过程定义 ID 和活动顺序 ID 的过程实例的一般信息, 并完成表的生成。所生成的活动分析表的属性成为数据库 40 的字段的一部分或全部, 与过程分析表相同。

图 5A 至 5D 的视图描述了预处理模块 34 的一个处理过程。

图 5A 示出了预处理模块被执行的初始屏幕。预处理模块通过按下处于图 5A 的屏幕顶端的 “DATA PREPROSESSING” 开关按钮而被启动。如图 5A 所示, 预处理模块包括一个用于由分析表提取的数据的基本表、和一个用于预处理过的数据的操作表。预处理方法依赖于一种将在后面用到的分析模块。因此根据分析的目标来执行一种恰当的预处理过程是很重要的。通过按下基本表上的 “Open DB” 按钮来执行分析表数据的载入模块, 由此而载入分析表的数据。图 5B 示出了载入的一种结果。基本表的属性信息窗口示出了属性的名称和类型。当用户点击了一个想要的属性, 基本表的属性信息窗口将示出属性的特性。属性的特性以符号格式示出了一个变量和频率, 以及数字格式的最大、最小、标准偏差和平均值。

为了去除不必要的属性，如图 5C 所示，在选项框中取消了对不必要属性的选择，并且通过按下“Apply Filter”开关按钮来执行属性选择和去除模块，由此而得到图 5D 的结果。也就是说，在去除不必要的属性之前属性的数目是 62，而在去除之后缩小到 35。另一方面，在预处理过程之后，另外需要一个通过符号变量来划分数字变量的过程来进行关联性分析。在此情况下，将使用离散化过滤器及两个算法：第一个是用来将一个最大值区域和一个最小值区域划分为属性分布中的预定区域并且离散化变量的方法，即无管理的离散化；第二个是用来根据类变量的分布来离散化属性的方法，即有管理的离散化。另一方面，记录过滤器是一个用于从分析对象中排除某个特定记录（意思与实例相同）的过滤器。例如，用记录过滤器来从分析对象中排除图 4C 中过程实例标识符 PROCID 在 120 之后的过程实例。另外，当一个变量是空值时使用遗漏值过滤器。主要地，遗漏值过滤器使用一种用于将数字变量替换为平均值，以及将符号变量替换为一个最常被生成的模式。

当预处理模块 34 处理完与去除属性、转换变量以及去除记录相关的预处理过程后，数据统计模块 36 为数据集提供文本或图形形式的基本的统计值以掌握变量的基本关联性。

图 6A 至 6C 的视图描述了数据统计模块 36 的一个处理过程。图 6A 示出了当数据统计模块 36 执行时的初始屏幕。通过按下图 6A 屏幕顶端的“DATA STATISTICS”开关按钮来启动数据统计模块。当用户点击了“DATA STATISTICS”按钮，由预处理数据装载模块所预处理的数据就被装载以为每一个变量提供统计值。也就是，对数字变量示出最小、最大、平均和标准偏差，对符号变量示出模式和频率。当用户点击了图 6A 右下端的“BASIC STATISTICS DISTRIBUTION”按钮时，示出模块将被执行以圆图的形式示出符号变量，以关于基本统计值的直方图形式示出数字变量。图 6B 示出了基本统计值的圆图和直方图。另外，当用户点击了图 6A 左下端的“SECONDARY SCATTER DIAGRAM”按钮时，将示出两个属性

之间的基本关联性。图 6C 说明了一个示出了表示创建时间 CREATIONTIME1 (a.m.和 p.m.格式) 的变量与表示延误过程实例 STATE1 的变量之间的关联性。可以看出, 当变量 CREATIONTIME1 的值为 AM, 并且变量 STATE1 的值为 NOOverDue 时, 这两个变量之间有非常紧密的关联性。

分析模块 38 在预处理模块 34 预处理的变量和属性以及数据的基础上, 分析决策树或关联性。

图 7A 至 7E 是描述了决策树分析模块的一个处理过程的视图。

决策树分析以决策树的形式为对象变量(类变量)产生可能的规则。

图 7A 示出了决策树分析模块执行时的初始屏幕。通过按下图 7A 顶端的“DECISION MAKING TREE ANALYSIS”开关按钮来开始决策树分析模块。当决策树分析模块被执行时, 由用户选择用于决策分析的算法。根据本发明, 由 J Ross Quinlan 所建议的 C4.5 算法被用作决策树分析算法, 但并不意味着仅限于此。如图 7B 所示, 当用户选择了算法, 决策树分析模块装载预处理的属性数据, 并且让用户决定是否通过变量来使用数据作分析。图 7C 示出了一个选择完分析用的变量以及对类完成指定后的状态。类指出了一个是可能规则的演绎对象的变量。在决策树分析中, 一对类变量以及一个特定的值成为规则的结果。例如, 当用户想要获得针对表示延误过程实例 STATE1 (变量 STATE1 和变量 OverDUE 的对) 的延误情况的可能规则, 变量 STATE1 成为类, 并且为了决策树分析, 此类需要是符号型的。因此, 在预处理过程中数字型的类变量必须要由离散化过滤器转化为符号变量。因为决策树分析是一种直接的分析方法, 表示分析对象的变量之一必须要被指定为类。

参考图 7C, 最后一个变量 STATE2 被选择不用作分析, 被指定为类的变量为 STATE1。当选择完了变量, 用户点击“EXECUTION”按钮来进行决策树分析。分析结果如图 7D 所示以文本的形式示出。通过点击“VISUALIZATION”按钮, 如图 7E

所示还可以以图形的形式示出。图 7D 和 7E 示出了当类变量是 STATE1 时的关于搜索影响延误过程的变量的分析。对 STATE1 最有影响作用的变量是 CREATIONDATE1，随后的影响变量是 AR10、CREATIONTIME3 和 AR5。

也就是说，过程在下面的场合下是延误的：

- 1)过程在下午执行，
- 2)过程在上午执行，并且第十活动的执行时间低于 5，
- 3)过程在上午执行，第十活动的执行时间高于 5，并且过程在 6 之后执行，
- 4)过程在上午执行，第十活动的执行时间高于 5，过程在 6 之后执行，并且第五活动的执行时间高于 1512。

如图 7E 所示，提供了四个内部节点。每一个节点示出为一个圆圈形。外部的圆圈示出了类变量在父节点中的分布率，内部圆圈示出了根据相应节点的标准间隔的类变量的间隔率。矩形表示一个叶节点，括号内的数字是属于相应叶节点的数据集中的实例的数目。当树很复杂时，可以通过整理来减少复杂性。特别地，通过点击某个特定节点，可以示出从最重要的节点到相应的节点所生成的规则。

在上面解释了通过使用经过预处理的数据来分析决策树的过程，但是另外一种分析可以用来分析关联性。

关联性分析意味着在支持和可靠性的基础上搜索数据库中的独立项之间的关联性或规则。这里，项是包括属性和属性的特定值的对，关联性分析是用于在所有在给定数据库中的可能的项的集合之间搜索关联性的一种分析方法。首先，支持意味着满足条件 X 和结果 Y 的过程实例或执行活动的数目与所有过程实例或执行活动的数目的比值。

$S(X,Y)=P(X,Y)/$ 所有过程实例或执行活动实例的数目

X: 条件

Y: 结果

P(X,Y): 过程或执行活动的实例（包括 X 和 Y 联合组成的三

项)的数目

$S(X,Y)$: 满足规则“如果条件是 X,则结果是 Y”

这种支持可以被认为是规则的统计重要性。因为支持意味着一个模型或规则经常被生成的频率,所以它应该被增加以提高模型或规则的应用。

另一方面,可靠性意味着满足结果 Y 的过程实例或执行活动的实例与满足条件 X 的过程实例或执行活动的实例的比值。这种比值成为测量规则“如果满足条件 X,则结果为 Y”的准确性的一个指标。因此,高准确性产生了准确的预测。

$C(X,Y)=P(X,Y)/P(X)$

X: 条件

Y: 结果

$P(X,Y)$: 包括由 X 和 Y 的联合组成的三个项的过程或执行活动的实例的数目

$C(X,Y)$: 规则“如果条件是 X,则结果是 Y”的可靠性

总之,规则的可靠性示出了结果可以被应用到条件中的频率,而支持示出了整个规则的可靠性。

另一方面,用于分析关联性的所有变量(作为分析对象)必须是符号型的。因此,数字变量需要通过预处理过程中的数字化过滤器被转化成符号变量。

图 8A 至 8D 是描述了关联性分析模块的一个处理过程的视图。

图 8A 示出了关联性分析模块被执行时的初始屏幕。关联性分析模块是通过按下图 8A 的屏幕的顶端的“ASSOCIATION”开关按钮而被启动的。当关联性模块执行时,用户需要选择用于分析的算法。根据本发明,Apriori 算法被用作分析关联性的算法,但并不仅限于此。如图 8B 所示,当用户选择了算法,关联性分析模块将提供一个用于设置分析用的输入变量的输入窗口。输入变量包括许多所需的规则、最小可靠性、最小支持边界值、支持的降低值、和最大支持边界值。当规则被构造后,将确定是否使用一个兴趣(interest)项,并

且用作兴趣项的一个变量和一个兴趣项值被指定以将兴趣项加入到规则的结果中。

当设置完在 Apriori 算法中执行分析模块所需的输入变量后，用户点击“ASSOCIATION ANALYSIS START”按钮以执行 Apriori 算法。关联性分析模块执行 Apriori 算法，并如图 8C 所示生成一个执行结果。这里，关联性分析结果被以文本的形式示出，但是可以通过点击“OUTPUT”按钮而如图 8D 所示以表的形式示出。

图 8C 和 8D 的分析结果是在图 8B 中输入的输入值的基础上生成的规则。图 8D 的分析结果中的规则 10 具有如下的意思：

条件 1) 过程是在星期五上午执行的 (CREATETIME1=AM, CREATIONTIME2=FRI)

条件 2) 第 2 活动的执行时间超过 179 (AR2='179-inf')

条件 3) 第 9 活动的执行时间低于 112 (AR9='-inf-112')

条件 4) 第 10 活动的执行时间低于 19 (AR9='-inf-19')

结论) 过程是延误的 (STATE1=OverDue)

另外，规则 10 的支持和可靠性分别是 19% 和 100%。因此，整个分析对象数据库的 19% 的过程可以被上述的规则所解释，并且规则的可靠性是 100%。也就是，当满足条件时，与结果相对应的结论总是会出现。

其它规则可以用与规则 10 相似的方式来解释。

图 9 示出了一个根据本发明的 workflow 挖掘方法的流程图。

首先，当 workflow 挖掘想要在一个特定的过程或活动上执行时，将生成一个对应的过程分析表或活动分析表 (S100)。

当生成完分析表后，预处理过程将作为用于决策树或关联性分析的在先过程而执行 (S110)。在预处理过程中，数据从分析表中装载，特定数据集的属性被去除，特定区域的记录被去除，或者通过适当的区域划分将数字变量转化为符号变量。特别是对关联性分析，数字变量必须被转化成符号变量。

此后，是一个用来确定选择性经过预处理的数据的基本统计资

料的过程（S120）。也就是，统计资料示出了数字变量的最大、最小、平均和标准偏差，以及与经过预处理的数据相关的符号变量的模式和频率。另外，如果需要，可以以直方图的形式示出数字变量，以圆图的形式示出符号变量。

当预处理过程结束后，可以用决策树分析或关联性分析来分析数据集。为了分析决策树，用于决策树分析的算法被选择（S130），预处理属性数据被装载（S140），是否使用数据来分析由变量来决定，并且类变量被指定（S150）。这里，类变量必须是符号变量。当数字变量被指定为类变量时，程序将转向预处理过程来完成转化。当变量的选择完成后，用于决策树分析的算法被执行来产生结果（S160）。

另一方面，当打算在预处理过程之后执行关联性分析时，用于关联性分析的算法被选择（S170）。当选择完算法后，用于设置分析关联性所需的输入变量的输入窗口将被提供（S180）。在输入窗口中指定了规则的数目、最小可靠性、最小支持边界值、支持的降低值、最大支持边界值、对兴趣项的使用、兴趣项变量和兴趣项值（S190）之后，用于关联性分析的算法将被执行来生成结果（S200）。

本发明在不脱离其精神和本质特征前提下，可以有多种具体实施方式，应当理解上述实施例并不仅限于上述的任何细节，而应该在所附权利要求所定义的精神和范围内被广泛地解释，因此所有落在权利要求的边界和范围内的或者与这些边界和范围等价的变化和修改都试图包含在附加权利要求内。

如前面所讨论的，根据本发明， workflow挖掘系统和方法通过在工作流系统的运行过程中累积的记录数据中以及在由连接到 workflow系统的应用系统所累积的数据中使用特定的数据发掘技术，以评价、分析和确定过程或活动的先前执行结果，从而系统性地重调整通用的商务过程。

图1

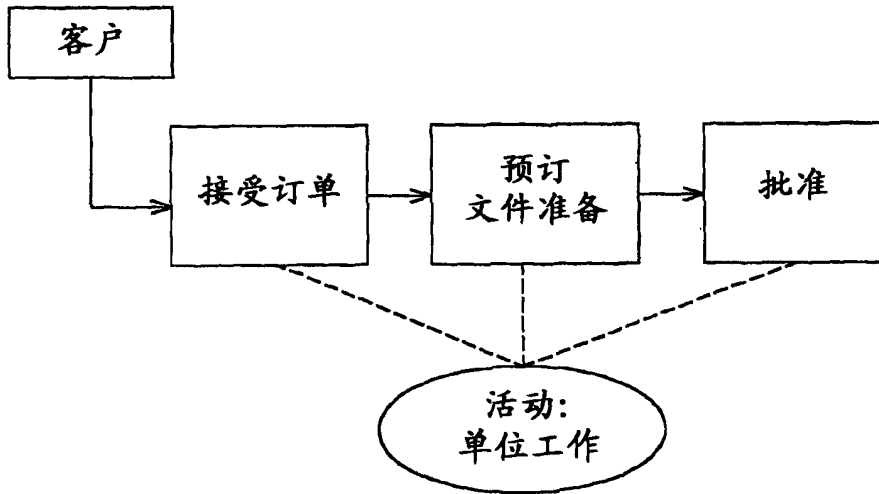
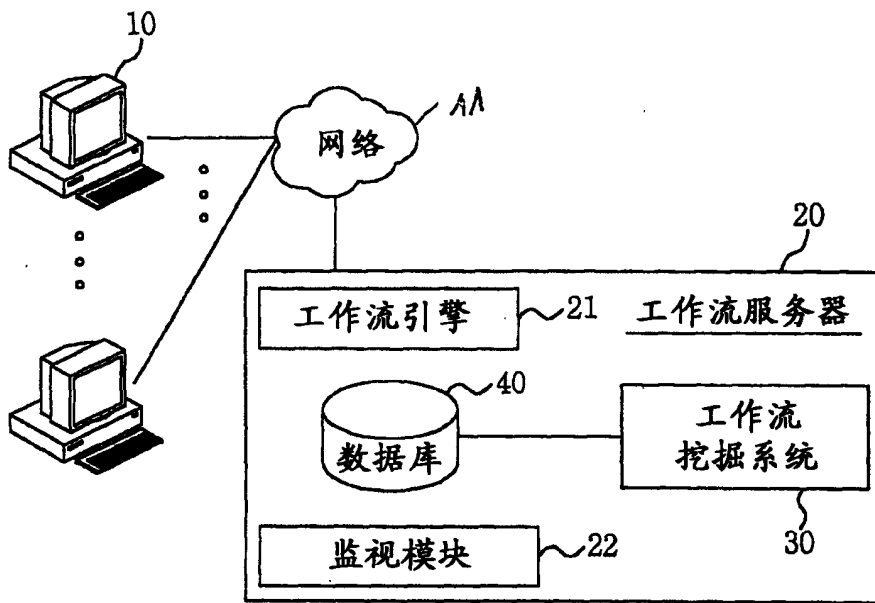


图2



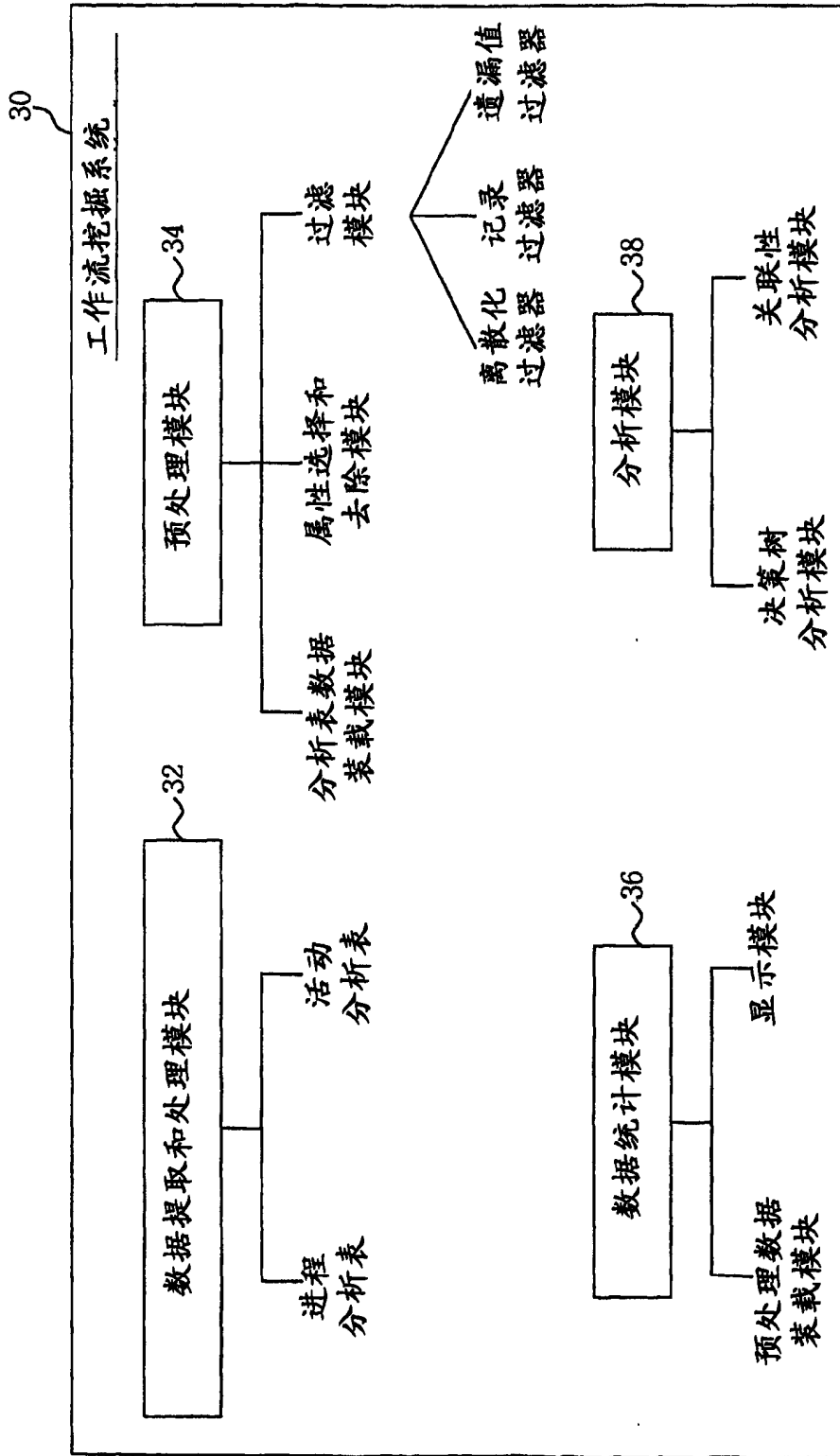


图 3

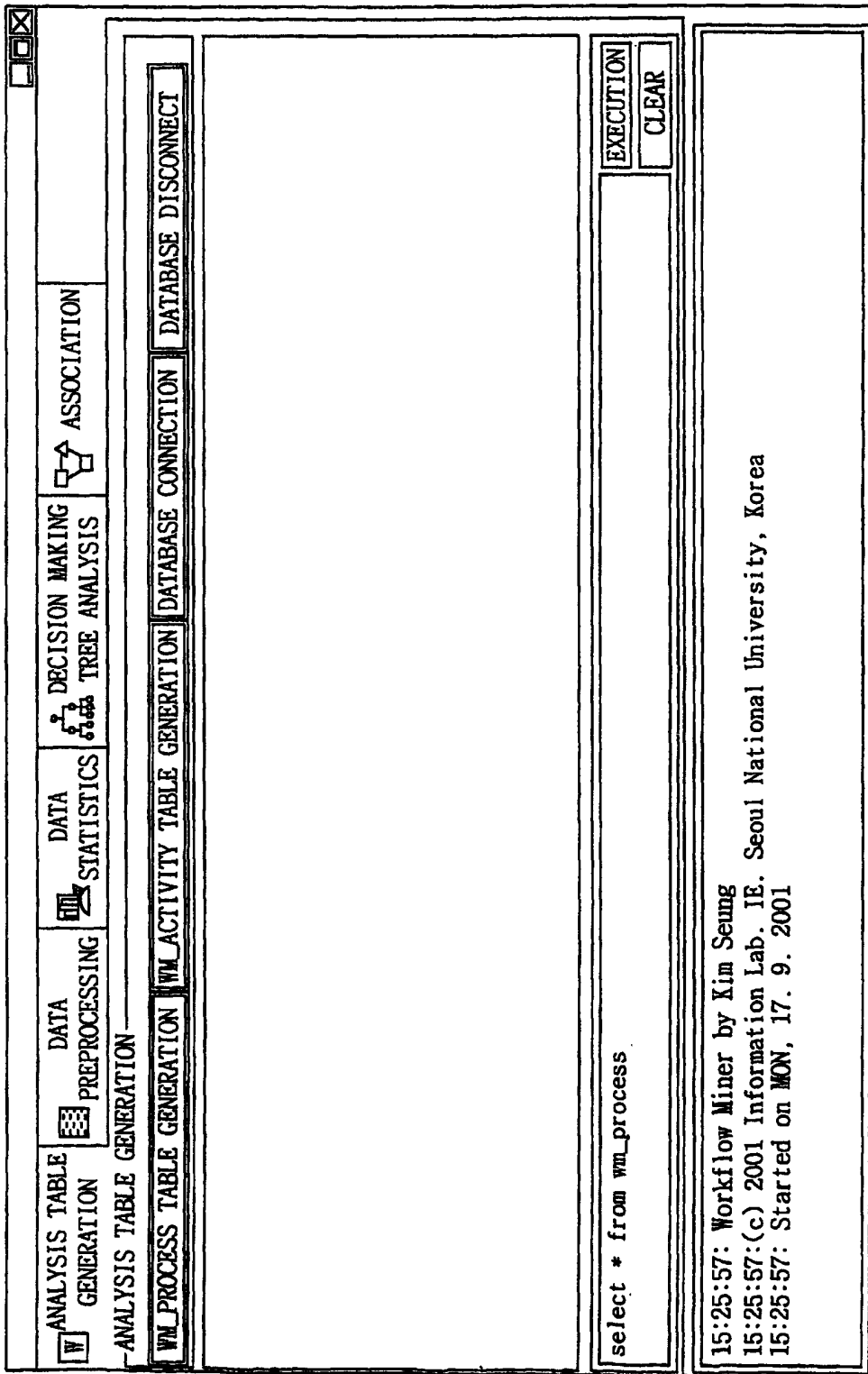
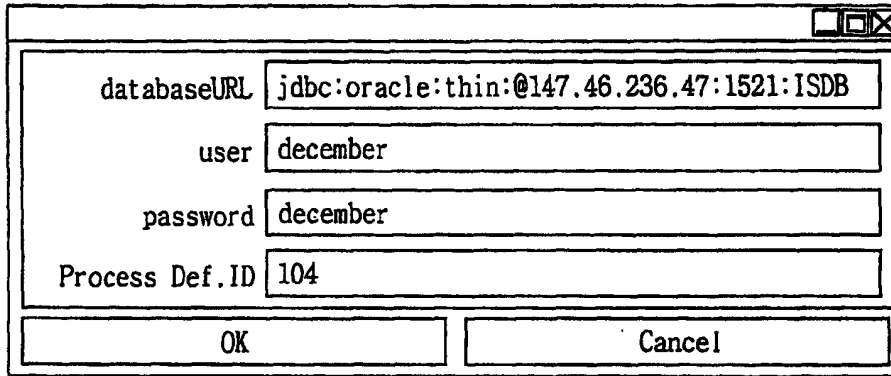


图 4A

图 4B



The image shows a dialog box with a title bar containing minimize, maximize, and close buttons. The dialog box contains four input fields with the following labels and values:

- databaseURL: jdbc:oracle:thin:@147.46.236.47:1521:ISDB
- user: december
- password: december
- Process Def.ID: 104

At the bottom of the dialog box, there are two buttons: "OK" and "Cancel".

ANALYSIS TABLE
GENERATION

DATA
PREPROCESSING

DATA
STATISTICS

DECISION MAKING
TREE ANALYSIS

ASSOCIATION

-ANALYSIS TABLE GENERATION

WM ACTIVITY TABLE GENERATION

DATABASE CONNECTION

DATABASE DISCONNECT

SVRID	PROCID	URGENT	PASSWDF	CNTCNT	ATTACHMT	NAME	CREATION	CREATION	CREATION
000000'001	108	T	F	0	0	Life Insurance Sales Process	AM	FRI	8
000000'001	109	F	F	0	0	Car Insurance Sales Process	AM	FRI	8
000000'001	110	F	F	0	0	Fire Insurance Sales Process	AM	FRI	8
000000'001	111	F	F	0	0	Fire Insurance Sales Process..	AM	FRI	8
000000'001	112	T	F	0	0	Car Insurance Sales Process2	AM	FRI	8
000000'001	113	F	F	0	0	Car Insurance Sales Process	AM	FRI	8
000000'001	114	T	F	0	0	Life Insurance Sales Process	AM	FRI	8
000000'001	115	F	F	0	0	Life Insurance Sales Process	AM	FRI	8
000000'001	116	F	F	0	0	Car Insurance Sales Process	AM	FRI	8
000000'001	117	F	F	0	0	Fire Insurance Sales Process	AM	FRI	8
000000'001	118	F	F	0	0	Life Insurance Sales Process	AM	FRI	8
000000'001	119	F	F	0	0	Life Insurance Sales Process	AM	FRI	9
000000'001	120	F	F	0	0	Car Insurance Sales Process	AM	FRI	9

select * from wm_process

EXECUTION

CLEAR

15:25:57: Workflow Miner by Kim Seung
 15:25:57:(c) 2001 Information Lab. IE. Seoul National University, Korea
 15:25:57: Started on MON, 17. 9. 2001
 15:25:57: DATABASE CONNECT OK.

图 4C

图 4D

The image shows a dialog box with a title bar containing minimize, maximize, and close buttons. The dialog box contains five input fields, each with a label to its left:

- databaseURL: jdbc:oracle:thin:@147.46.235.190:1521:ISDB
- user: december
- password: december
- Process Def. ID: 104
- Activity Seq. ID: 8

At the bottom of the dialog box, there are two buttons: "OK" on the left and "Cancel" on the right.

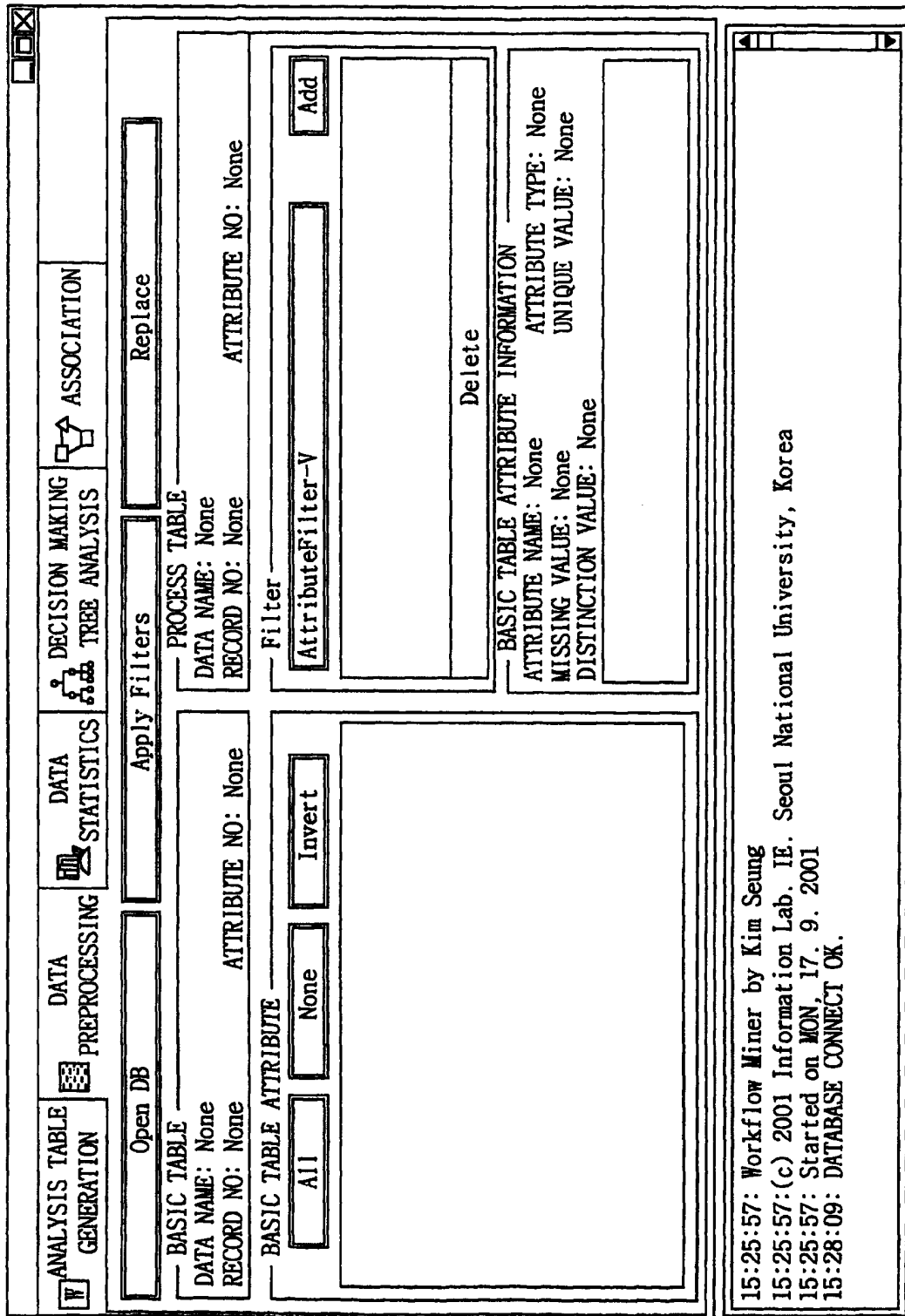


图 5A

ANALYSIS TABLE
GENERATION

DATA
PREPROCESSING

DATA
STATISTICS

DECISION MAKING
TREE ANALYSIS

ASSOCIATION

Open DB

Apply Filters

Replace

BASIC TABLE
DATA NAME: QueryResult
RECORD NO: 27 ATTRIBUTE NO: 62

BASIC TABLE ATTRIBUTE
All None Invert

No.	ATTRIBUTE NAME	DATA TYPE	DATA TYPE CHANGE
1	SVRID	SYMBOL	
2	PROCID	NUMERICAL VALUE	STANDARD VALUE
3	URGENT	SYMBOL	
4	PASSWDFLAG	SYMBOL	
5	CMTCNT	NUMERICAL VALUE	STANDARD VALUE
6	ATTACHCNT	NUMERICAL VALUE	STANDARD VALUE
7	NAME	SYMBOL	
8	CREATIONTIME1	SYMBOL	
9	CREATIONTIME2	SYMBOL	
10	CREATIONTIME3	NUMERICAL VALUE	STANDARD VALUE
11	CREATOR	SYMBOL	

PROCESS TABLE
DATA NAME: QueryResult
RECORD NO: 27 ATTRIBUTE NO: 62

Filter: AttributeFilter-V Add

Delete

BASIC TABLE ATTRIBUTE INFORMATION
ATTRIBUTE NAME: None ATTRIBUTE TYPE: None
MISSING VALUE: None UNIQUE VALUE: None
DISTINCTION VALUE: None

15:25:57: Started on MON, 17. 9. 2001
 15:28:09: DATABASE CONNETOK.
 15:30:10: BASICTABLE load : QueryResult (27 Instances)
 15:30:10: PROCESSING TABLE load : QueryResult (27 Instances)

图 5B

ANALYSIS TABLE
GENERATION

DATA
PREPROCESSING

DATA
STATISTICS

DECISION MAKING
TREE ANALYSIS

ASSOCIATION

Open DB

Apply Filters

Replace

BASIC TABLE

DATA NAME: QueryResult

RECORD NO: 27

PROCESS TABLE

DATA NAME: QueryResult

RECORD NO: 27

ATTRIBUTE NO: 62

All

None

Invert

No.	ATTRIBUTE NAME	DATA TYPE	DATA TYPE CHANGE
1	SVRID	SYMBOL	
2	PROCID	NUMERICAL VALUE	STANDARD VALUE
3	URGENT	SYMBOL	
4	PASSWFLAG	SYMBOL	
5	CONTONT	NUMERICAL VALUE	STANDARD VALUE
6	ATTACHCNT	NUMERICAL VALUE	STANDARD VALUE
7	NAME	SYMBOL	
8	CREATIONTIME1	SYMBOL	
9	CREATIONTIME2	SYMBOL	
10	CREATIONTIME3	NUMERICAL VALUE	STANDARD VALUE
11	CREATOR	SYMBOL	

Filter: AttributeFilter-V

Delete

BASIC TABLE ATTRIBUTE INFORMATION

ATTRIBUTE NAME: ACTYPE14 ATTRIBUTE TYPE: SYMBOL

MISSING VALUE: 0(0%) UNIQUE VALUE: 0(0%)

DISTINCTION VALUE: 0

15:25:57: Started on MON, 17. 9. 2001

15:28:09: DATABASE CONNETOK.

15:30:10: BASICTABLE load : QueryResult (27 Instances)

15:30:10: PROCESSING TABLE load : QueryResult (27 Instances)

图 5C

ANALYSIS TABLE
GENERATION

DATA
PREPROCESSING

DATA
STATISTICS

DECISION MAKING
TREE ANALYSIS

ASSOCIATION

Open DB

Apply Filters

Replace

BASIC TABLE

DATA NAME: QueryResult

RECORD NO: 27 ATTRIBUTE NO: 62

PROCESS TABLE

DATA NAME: QueryResult

RECORD NO: 27 ATTRIBUTE NO: 62

BASIC TABLE ATTRIBUTE

All None Invert

No.	ATTRIBUTE NAME	DATA TYPE	DATA TYPE CHANGE
21	ACTNAME6	SYMBOL	
22	ACTTYPE6	SYMBOL	
23	AR7	NUMERICAL VALUE	STANDARD VALUE
24	ACTNAME7	SYMBOL	
25	ACTTYPE7	SYMBOL	
26	AR8	NUMERICAL VALUE	STANDARD VALUE
27	ACTNAME8	SYMBOL	
28	ACTTYPE8	SYMBOL	
29	AR9	NUMERICAL VALUE	STANDARD VALUE
30	ACTNAME9	SYMBOL	
31	ACTTYPE9	SYMBOL	

Filter

AttributeFilter-V

Add

Delete

BASIC TABLE ATTRIBUTE INFORMATION

ATTRIBUTE NAME: None ATTRIBUTE TYPE: None

MISSING VALUE: None UNIQUE VALUE: None

DISTINCTION VALUE: None

17:03:16: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62(27 instanc

17:03:17: BASIC TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62(27 instances)

17:03:24: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.fil

17:03:26: BASIC TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filters

图 5D

ANALYSIS TABLE GENERATION

DATA PREPROCESSING

STATISTICS

DECISION MAKING

TREE ANALYSIS

ASSOCIATION

BASIC STATISTICS

DATA NAME: QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filters.AttributeFilter

RECORD NO: 27 ATTRIBUTE NO: 11

(1) URGENT
MODE: F FREQUENCY: 19

(2) NAME
MODE: Car Insurance Sales Process FREQUENCY: 7

(3) CREATIONTIME1
MODE: AM FREQUENCY: 24

(4) CREATIONTIME2
MODE: FRI FREQUENCY: 25

(5) CREATORTNAME
MODE: Administrator FREQUENCY: 23

(6) DEPTID
MODE: 9000000000 FREQUENCY: 23

INPUT ATTRIBUTE SELECTION

No.	ATTRIBUTE NAME	DATA TYPE
1	<input type="checkbox"/> URGENT	SYMBOL
2	<input type="checkbox"/> NAME	SYMBOL
3	<input type="checkbox"/> CREATIONTIME1	SYMBOL
4	<input type="checkbox"/> CREATIONTIME2	SYMBOL
5	<input type="checkbox"/> CREATORTNAME	SYMBOL
6	<input type="checkbox"/> DEPTID	SYMBOL
7	<input type="checkbox"/> AR2	SYMBOL
8	<input type="checkbox"/> AR5	SYMBOL
9	<input type="checkbox"/> AR9	SYMBOL
10	<input type="checkbox"/> AR10	SYMBOL
11	<input type="checkbox"/> STATE1	SYMBOL

17:05:40: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filte

17:05:41: BASIC TABLE load : QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filter-

17:09:31: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filte

17:09:32: BASIC TABLE load : QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62-wminer.filter-

图 6A

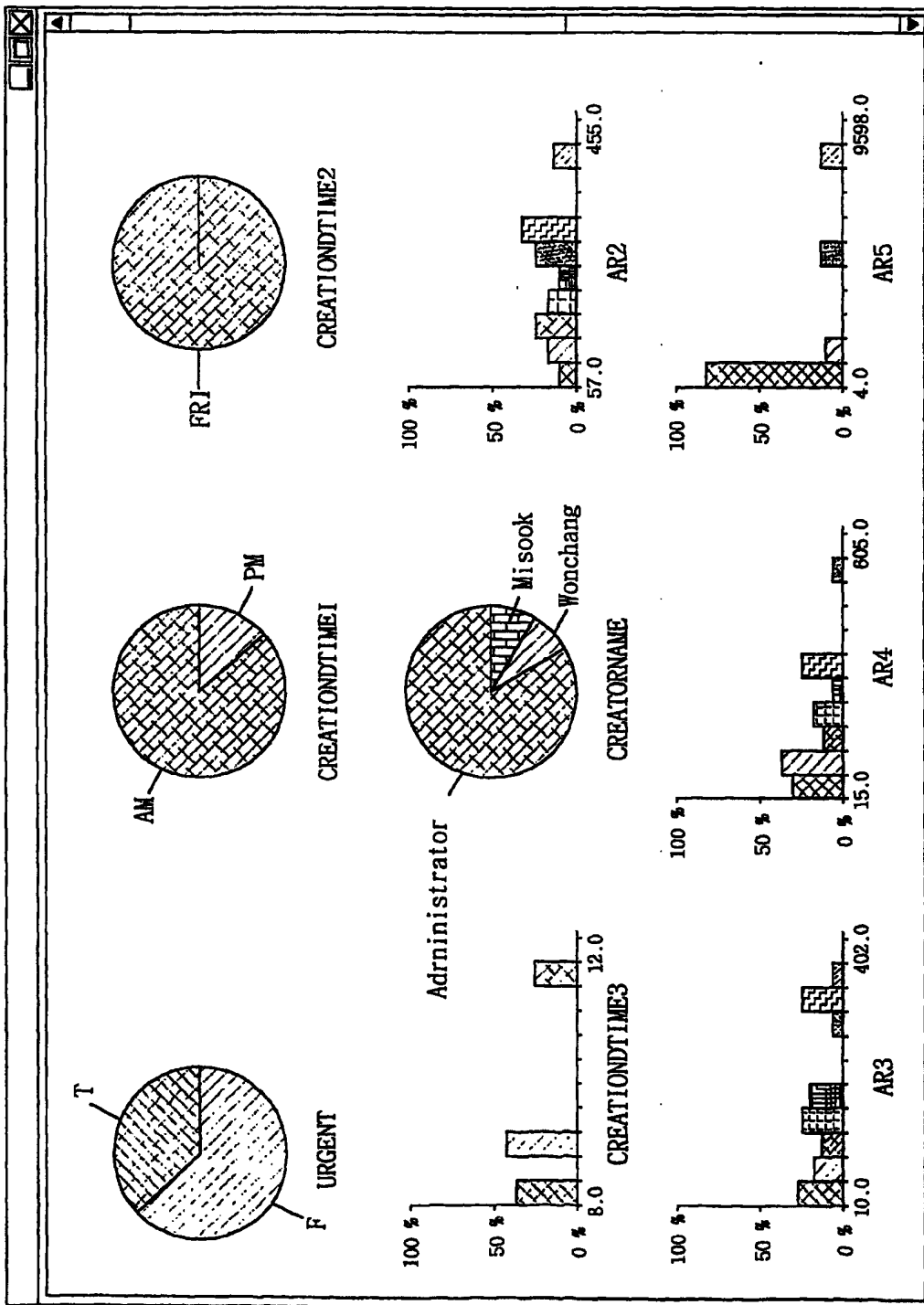
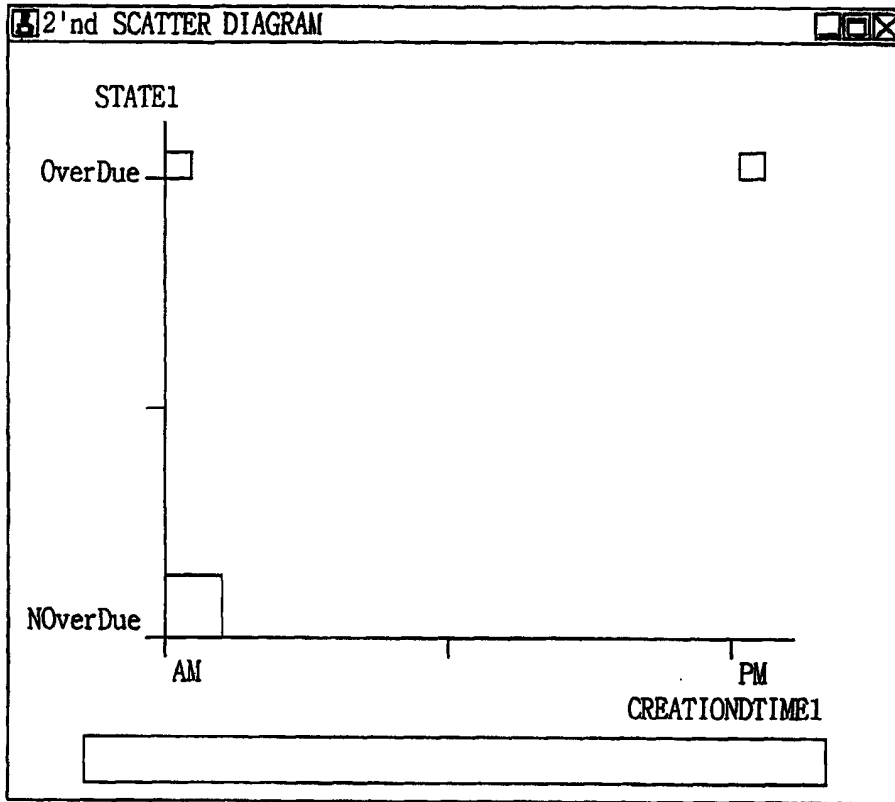


图 6B

图 6C



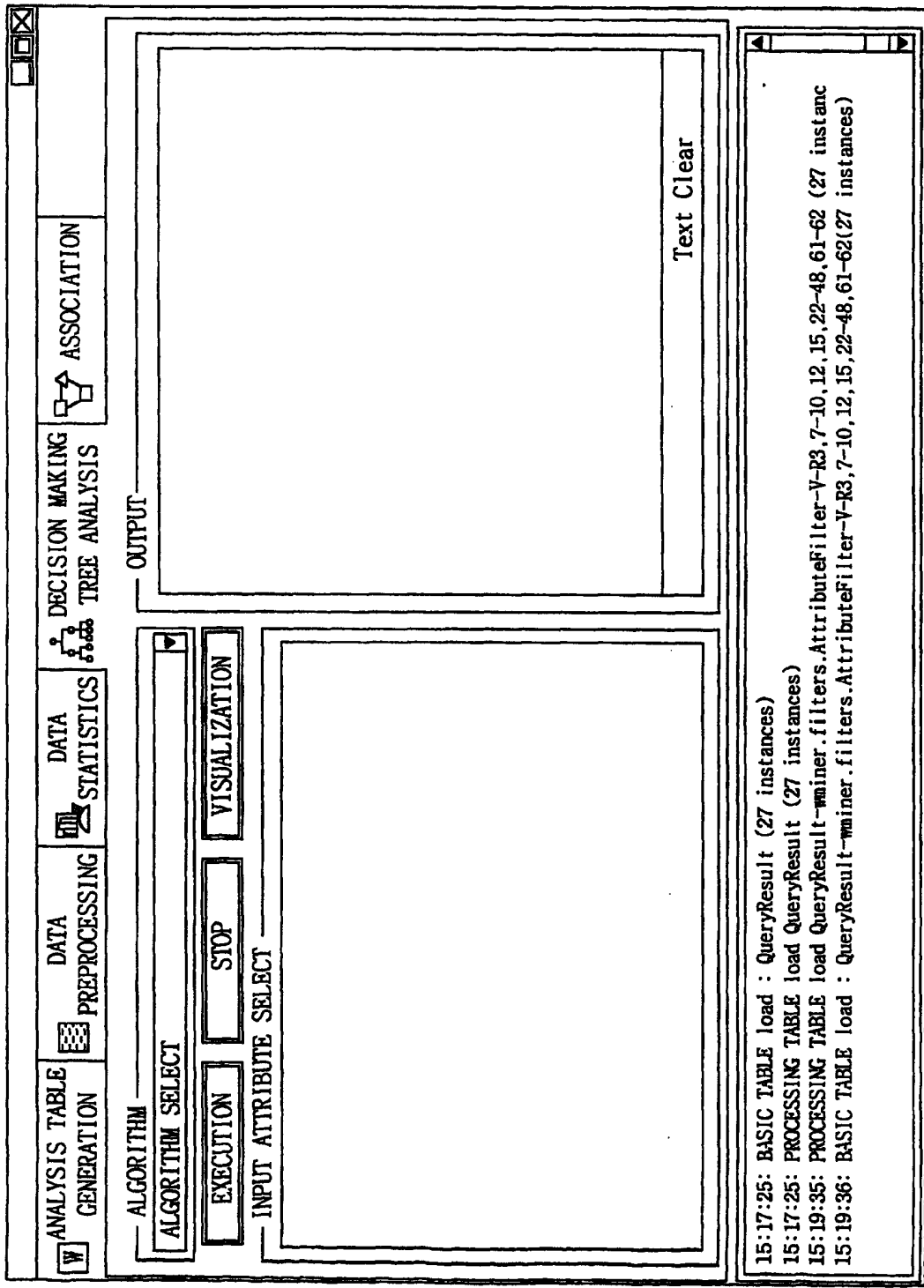


图 7A

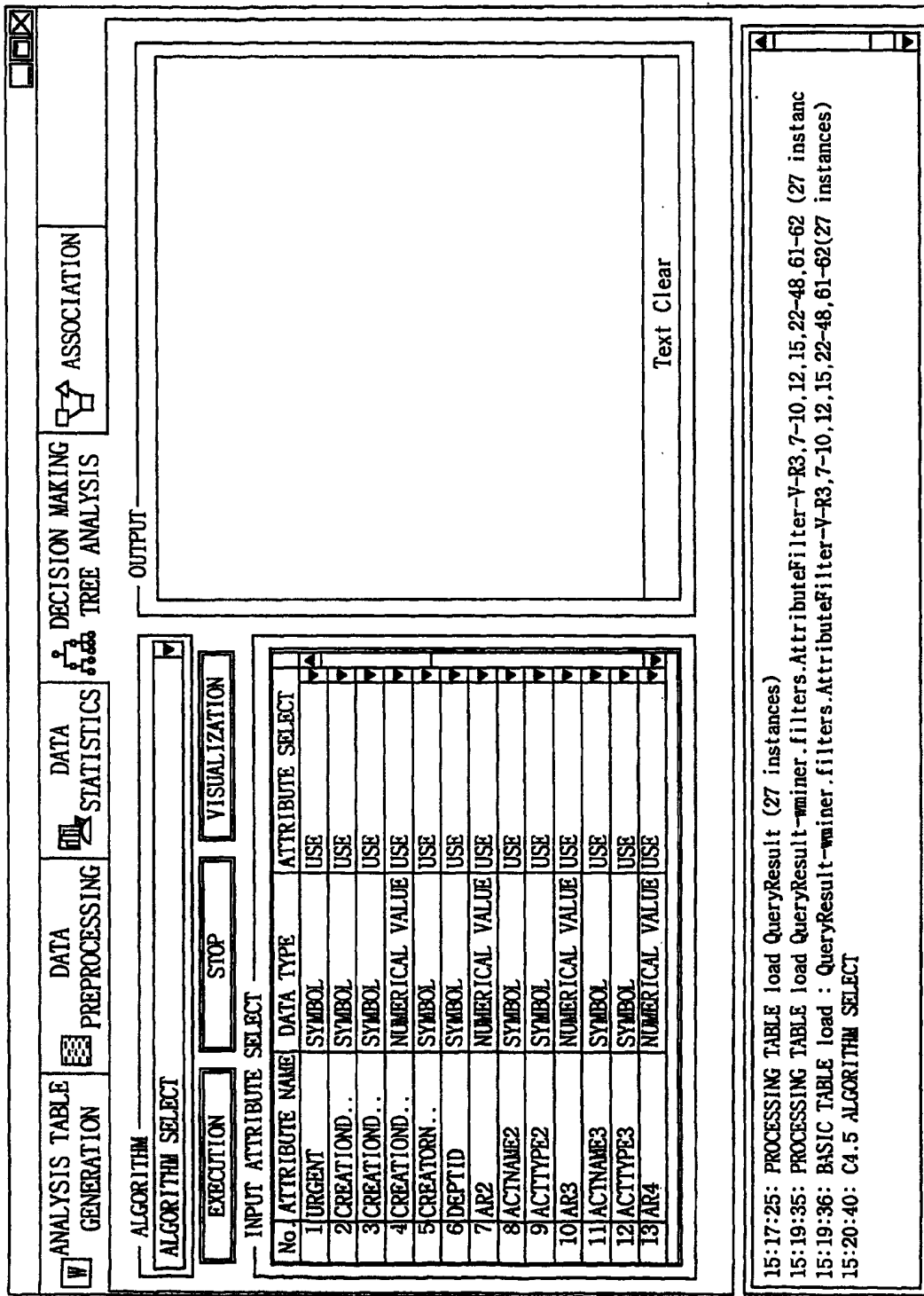


图 7B

ANALYSIS TABLE GENERATION

DATA PREPROCESSING

DATA STATISTICS

DECISION MAKING TREE ANALYSIS

ASSOCIATION

ALGORITHM

ALGORITHM SELECT

EXECUTION

STOP

VISUALIZATION

INPUT ATTRIBUTE SELECT

No.	ATTRIBUTE NAME	DATA TYPE	ATTRIBUTE SELECT
15	ACTYPE4	SYMBOL	USE
16	AR5	NUMERICAL VALUE	USE
17	ACTNAME5	SYMBOL	USE
18	ACTYPE5	NUMERICAL VALUE	USE
19	AR6	NUMERICAL VALUE	USE
20	ACTNAME6	SYMBOL	USE
21	ACTYPE6	SYMBOL	USE
22	AR7	NUMERICAL VALUE	USE
23	ACTNAME7	SYMBOL	USE
24	ACTYPE7	SYMBOL	USE
25	AR8	NUMERICAL VALUE	USE
26	ACTNAME8	SYMBOL	USE
27	STATE2	NUMERICAL VALUE	UNUSED

OUTPUT

Text Clear

15:17:25: PROCESSING TABLE load QueryResult (27 instances)

15:19:35: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62 (27 instances)

15:19:36: BASIC TABLE load : QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62(27 instances)

15:20:40: C4.5 ALGORITHM SELECT

图 7C

ANALYSIS TABLE
GENERATION

DATA
PREPROCESSING

DATA
STATISTICS

DECISION MAKING
TREE ANALYSIS

ASSOCIATION

ALGORITHM
C4.5

EXECUTION STOP VISUALIZATION

INPUT ATTRIBUTE SELECT

No.	ATTRIBUTE NAME	DATA TYPE	ATTRIBUTE SELECT
15	ACTTYPE4	SYMBOL	USE
16	AR5	NUMERICAL VALUE	USE
17	ACTNAME5	SYMBOL	USE
18	ACTTYPE5	NUMERICAL VALUE	USE
19	AR6	NUMERICAL VALUE	USE
20	ACTNAME6	SYMBOL	USE
21	ACTTYPE6	SYMBOL	USE
22	AR7	NUMERICAL VALUE	USE
23	ACTNAME7	SYMBOL	USE
24	ACTTYPE7	SYMBOL	USE
25	AR8	NUMERICAL VALUE	USE
26	ACTNAME8	SYMBOL	USE
27	STATE2	NUMERICAL VALUE	UNUSED

OUTPUT

```

AR8
ACTNAME8
ACTTYPE8
AR9
ACTNAME9
ACTTYPE9
AR10
ACTNAME10
ACTTYPE10
(4) CLASS NAME : STATE1
C4.5 TREE
-----
CREATIONTIME1 = AM
| AR10 <= 5 THAN OverDue (4.0)
| AR10 > 5
| CREATIONTIME3 <= 6 THAN OverDue (1.0)
| CREATIONTIME3 > 6
| AR5 <= 1512 THAN OverDue (18.0)
Text Clear
                
```

15:19:36: BASIC TABLE load : QueryResult-wminer.filters.AttributeFilter-V-R3,7-10,12,15,22-48,61-62(27 instances).
15:20:40: C4.5 ALGORITHM SELECT
15:21:51: C4.5 ALGORITHM EXECUTION START
15:21:52: C4.5 ALGORITHM EXECUTION END

图 7D

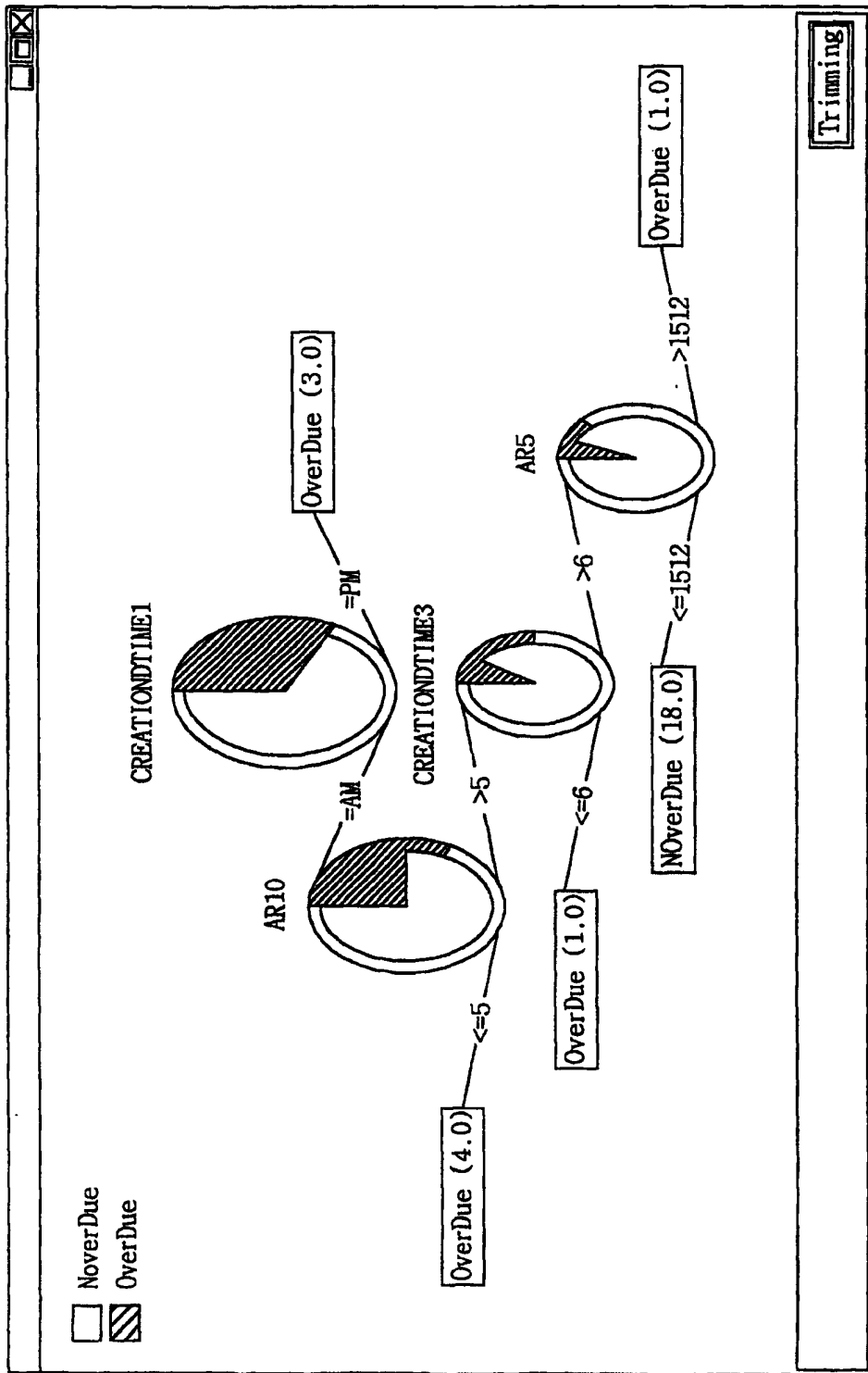


图 7E

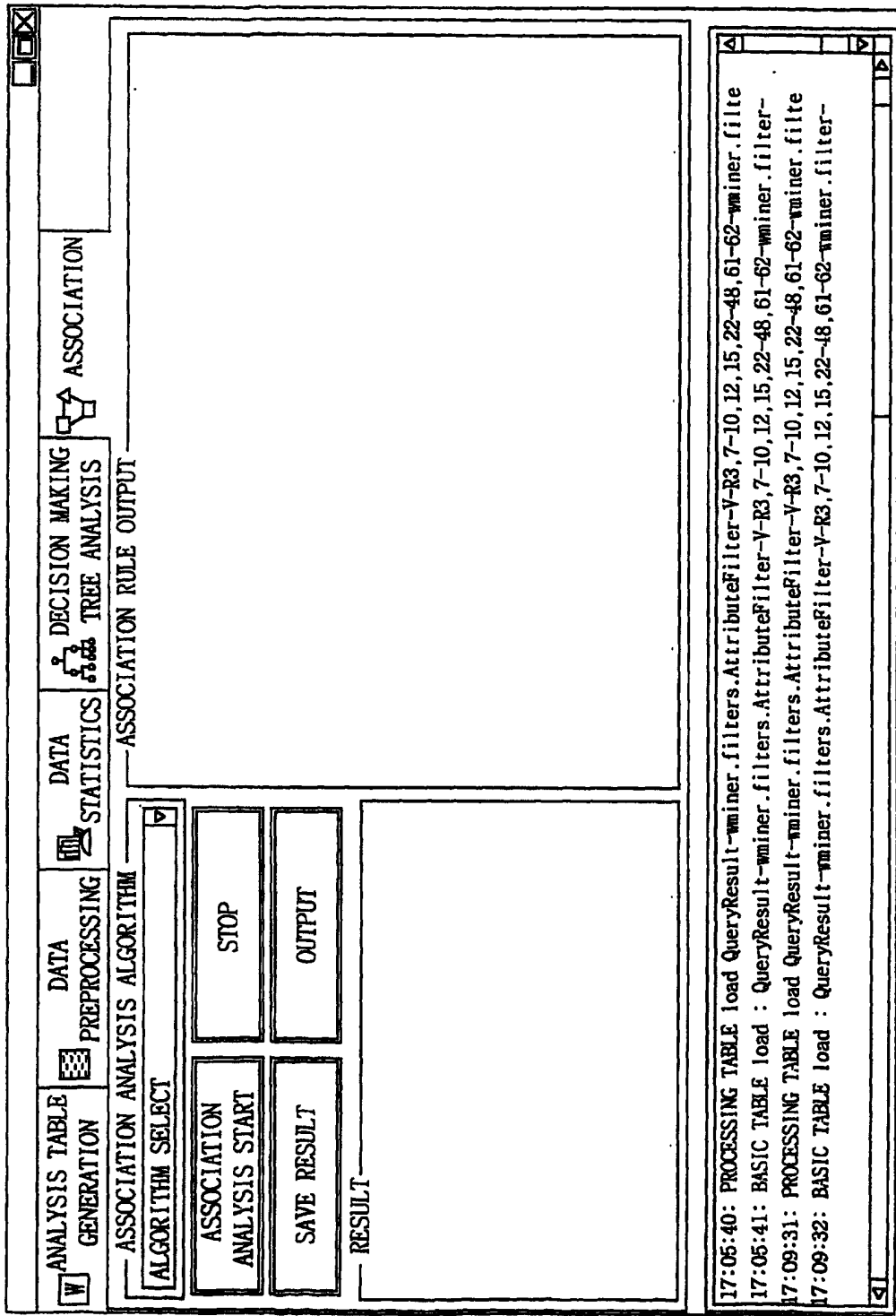


图 8A

图 8B

The dialog box is titled "ALGORITHM INPUT VALUE" and "INTEREST ITEM INPUT". It contains the following fields and controls:

- ALGORITHM INPUT VALUE:**
 - RULE NO: 10
 - MINIMUM RELIABILITY: 0.9
 - MINIMUM SUPPORT BOUNDARY VALUE: 0.1
 - SUPPORT DECREASE VALUE: 0.05
 - MAXIMUM SUPPORT BOUNDARY VALUE: 1
- INTEREST ITEM INPUT:**
 - INTEREST ITEM USE: USE
 - INTEREST ITEM: (Nom)STATE1
 - INTEREST VALUE: OverDue

Buttons: OK, Cancel

The screenshot shows a software window with several tabs at the top: ANALYSIS TABLE GENERATION, DATA PREPROCESSING, DATA STATISTICS, DECISION MAKING, and ASSOCIATION. The main area is divided into sections for algorithm selection and results display.

ASSOCIATION ANALYSIS ALGORITHM

- ALGORITHM SELECT
- ASSOCIATION ANALYSIS START
- SAVE RESULT
- STOP
- OUTPUT

ASSOCIATION RULE OUTPUT

- Large itemsets => L(3): 313
- Large itemsets => L(4): 512
- Large itemsets => L(5): 521
- Large itemsets => L(6): 333
- Large itemsets => L(7): 128
- Large itemsets => L(8): 26
- Large itemsets => L(9): 2

MOST BENEFICIAL RULE

- [CREATIONDTIME1] = AM [CREATORNAME] = Administrator [AR2] ='(179-int)'
- [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [DEPTID] = 900000000 [AR2]
- [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [CREATORNAME] = Administrato
- [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [AR2] = '(179-inf)' [AR9] =
- [CREATIONDTIME1] = AM [CREATORNAME] = Administrator [AR2] ='(179-int)'
- [CREATIONDTIME1] = AM [CREATORNAME] = Administrator [AR2] ='(179-int)'

RESULT

```

17:09:31: PROCESSING TABLE load QueryResult-wminer.filters.AttributeFilter-V-R3.7-10.12.15.22-48.61-62-wminer.filte
17:09:32: BASIC TABLE load : QueryResult-wminer.filters.AttributeFilter-V-R3.7-10.12.15.22-48.61-62-wminer.filter-
17:14:03: Apriori Algorithm Execution Start
17:14:06: Apriori Algorithm Execution End

```

图 8C

ASSOCIATION RULE WINDOW			
SUPPORTRATE(%)	RELIABILITY RATE(%)	RULE CONDITION	
1	C.19	1. [CREATIONDTIME1] = AM [CREATORNAME] = Administrat	⇒ [CEPTIC] = 900000000
2	C.19	2. [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [DEPT1]	⇒ [CREATORNAME] = AcmIn
3	C.19	3. [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [CREAT	⇒ [STATE1] = OverDue
4	C.19	4. [CREATIONDTIME1] = AM [CREATORNAME2] = FRI [AR2]	⇒ [CREATORNAME] = AcmIn
5	C.19	5. [CREATIONDTIME1] = AM [CREATORNAME] = Administrat	⇒ [CREATORNAME2] = FRI
6	C.19	6. [CREATIONDTIME1] = AM [CREATORNAME] = Administrat	⇒ [STATE1] = OverDue
7	C.19	7. [CREATIONDTIME1] = AM [DEPTID] = 900000000 [AR2]	⇒ [STATE1] = OverDue
8	C.19	8. [URGENT] = F [AR2] = '(179-inf)' [AR9] = '(-inf-11	⇒ [CREATORNAME1] = AM
9	C.19	9. [CREATIONDTIME1] = AM [CREATORNAME] = Administrat	⇒ [CREATORNAME2] = FRI
10	C.19	10. [CREATIONDTIME1] = AM [CREATIONDTIME2] = FRI [AR	⇒ [STATE1] = OverDue

图 8D

图9

