



[12] 发明专利说明书

专利号 ZL 200710000717.6

[45] 授权公告日 2009年3月18日

[11] 授权公告号 CN 100470547C

[22] 申请日 2007.1.10

[21] 申请号 200710000717.6

[73] 专利权人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

[72] 发明人 简宋全

[56] 参考文献

US2002/0128997A1 2002.9.12

US2002/0138492A1 2002.9.26

数据挖掘 - 电信客户流失分析预测. 王维佳, 缪柏其, 魏国省. 数理统计与管理, 第25卷第4期. 2006

数据挖掘技术在个人信用评估中的应用与实现. 刘利红, 韦薇. 华南金融电脑, 第3期. 2004

审查员 胡雅娟

[74] 专利代理机构 北京德琦知识产权代理有限公司

代理人 王琦 王诚华

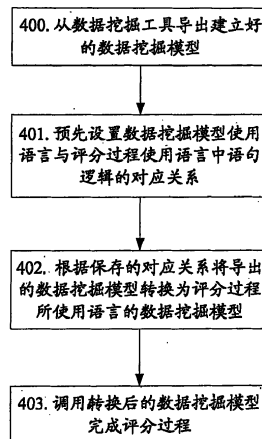
权利要求书4页 说明书18页 附图7页

[54] 发明名称

实现数据挖掘模型转换和应用的方法、系统及装置

[57] 摘要

本发明公开了一种实现数据挖掘模型应用的方法, 该方法包括: 导出建立好的数据挖掘模型, 根据预设的数据挖掘模型所使用语言与评分过程使用语言中语句逻辑的对应关系, 将数据挖掘模型转换为评分过程使用语言的数据挖掘模型; 将转换后的数据挖掘模型保存为存储过程或函数; 将评分数据作为所述存储过程或函数的参数, 通过调用所述存储过程或函数来完成评分过程。 本发明还公开了一种实现数据挖掘模型应用的系统和装置。 使用本发明使得数据挖掘模型的评分过程脱离数据挖掘工具和人力介入, 通过调用转换后的数据挖掘模型完成评分过程, 实现了数据挖掘模型的自动化应用。



1、一种实现数据挖掘模型应用的方法，其特征在于，该方法包括：

导出建立好的数据挖掘模型；

根据预设的数据挖掘模型所使用语言与评分过程使用语言中语句逻辑的对应关系，将数据挖掘模型转换为评分过程使用语言的数据挖掘模型；

将转换后的数据挖掘模型保存为存储过程或函数；

将评分数据作为所述存储过程或函数的参数，通过调用所述存储过程或函数来完成评分过程。

2、如权利要求1所述的方法，其特征在于，所述将数据挖掘模型转换为评分过程使用语言的数据挖掘模型的方法为：

搜索所述数据挖掘模型，当搜索到标识数据挖掘模型结束的字符串时，从所述标识数据挖掘模型结束的字符串所在行逐行向前搜索；

当搜索到转行符时，将转行符连接的两行合并；

当搜索到条件语句时，替换成评分过程使用语言的条件语句；

当搜索到数据挖掘工具根据目标字段生成的字符串时，则将字符串删去，输出字符串后的常数数值；

搜索到标识数据挖掘模型开始的字符串，则结束模型转换流程。

3、如权利要求1所述的方法，其特征在于，所述将数据挖掘模型转换为评分过程使用语言的数据挖掘模型的方法包括：

搜索所述数据挖掘模型，当搜索到标识数据挖掘模型开始的字符串时，从所述标识数据挖掘模型开始的字符串所在行逐行向后搜索；

当搜索到转行符时，将转行符连接的两行合并；

当搜索到条件语句时，替换为评分过程使用语言的条件语句；

当搜索到数据挖掘工具根据目标字段生成的字符串，则将字符串删去，输出字符串后的常数数值；

当搜索到标识数据挖掘模型结束的字符串，则结束模型转换流程。

4、如权利要求 1 所述的方法，其特征在于，所述将数据挖掘模型转换为评分过程使用语言的数据挖掘模型包括：

搜索所述数据挖掘模型，搜索到模型中标识细分分组的字符串，生成与分组个数相同的分组数组，从模型中标识细分分组的字符串所在行逐行向前搜索；

搜索到条件语句，利用条件语句中的常数数值计算出评分数据中每个字段的数据到各个分组中心的距离，输出所述距离到所述分组数组；

搜索到模型中标识细分分组结束的字符串，则结束搜索；

比较每个评分数据到各个分组中心的距离，并将所述评分数据归到距离最近的分组中。

5、如权利要求 4 所述的方法，其特征在于，所述利用条件语句中的常数数值计算出评分数据中每个字段的数据到各个分组中心的距离为：

利用所述条件语句后面第一个语句中的常数和各个分组在各字段上的中心值计算出评分数据中每个字段的数据到各个分组中心的距离。

6、如权利要求 4 所述的方法，其特征在于，所述搜索到模型中标识细分分组结束的字符串之前，该方法进一步包括：

从所述数据挖掘模型中搜索到根据离散属性变量生成的字符串，则搜索到该离散属性变量，将所述根据离散属性变量生成的字符串替换为该变量。

7、一种实现数据挖掘模型应用的系统，其特征在于，该系统包括模型建立导出装置、模型转换装置和评分装置，

所述模型建立导出装置，用于建立数据挖掘模型，将建立好的数据挖掘模型导出到模型转换装置；

所述模型转换装置，用于保存预先设置的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系；接收模型建立导出装置导出的数据挖掘模型，根据所述对应关系将所述数据挖掘模型转换为评分过程使用语言的数据挖掘模型，将转换后的数据挖掘模型保存为存储过程或函数，并输出到评分装置；

所述评分装置，用于接收所述模型转换装置输出的所述存储过程或函

数，通过调用所述存储过程或函数来完成评分过程。

8、如权利要求7所述的系统，其特征在于，所述模型转换装置包括搜索模块、对应关系存储模块和传输模块，

所述传输模块，用于接收模型建立导出装置导出的数据挖掘模型，传输到搜索模块；接收搜索模块传输的转换后的数据挖掘模型，将所述转换后的数据挖掘模型保存为存储过程或函数，并传输到评分装置；

所述搜索模块，用于接收传输模块传输的导出的数据挖掘模型，根据对应关系存储模块中保存的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系，将所述模型建立导出装置导出的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型，将转换后的数据挖掘模型传输到传输模块；

所述对应关系存储模块，用于保存预先设置的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系。

9、如权利要求7所述的系统，其特征在于，所述评分装置包括评分模块和调用模块，

所述调用模块，用于接收并保存模型转换装置传输的所述存储过程或函数，将所述转换后的数据挖掘模型传输到评分模块；

所述评分模块，用于接收调用模块传输的所述存储过程或函数，并通过调用所述存储过程或函数来完成评分过程。

10、一种模型转换装置，其特征在于，该装置包括搜索模块、对应关系存储模块和传输模块，

所述传输模块，用于接收模型建立导出装置导出的数据挖掘模型，传输到搜索模块；接收搜索模块传输的转换后的数据挖掘模型，将所述转换后的数据挖掘模型保存为存储过程或函数，并传输到评分装置；

所述搜索模块，用于接收传输模块传输的导出的数据挖掘模型，根据对应关系存储模块中保存的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系，将所述模型建立导出装置导出的数据挖掘模

型转换为评分过程所使用语言的数据挖掘模型,将转换后的数据挖掘模型传输到传输模块;

所述对应关系存储模块,用于保存预先设置的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系。

实现数据挖掘模型转换和应用的方法、系统及装置

技术领域

本发明涉及数据挖掘技术，具体涉及一种实现数据挖掘模型转换和应用的方法、系统及装置。

背景技术

随着数据库技术和数据库管理系统的迅速发展，数据的积累越来越多。激增的数据背后隐藏着许多重要的信息，目前的数据库系统可以高效地实现数据录入、查询和统计等功能，但无法发现数据中存在的关系和规则，无法根据现有的数据预测未来的发展趋势。缺乏挖掘数据背后隐藏的知识的手段，导致了“数据爆炸”但知识贫乏的现象。

数据库知识发现（KDD, Knowledge Discovery In Database）是人工智能、机器学习与数据库技术等多种学科相结合的产物，是从大量数据中提取出可信、新颖、有用并能被人理解的模式的高级处理过程。这里的模式就是知识，或者说是隐藏在数据背后的规律、关系或规则。

图 1 所示为现有技术 KDD 处理过程，如图 1 所示，KDD 处理过程主要包括数据选择、数据预处理、数据转换、数据挖掘和模式解释/知识评价五个步骤。数据挖掘（DM, Data Mining）是 KDD 中的一个重要步骤，用于发现转换后的数据中存在的关系和规则，下文用数据挖掘来指代 KDD 的整个分析过程。

KDD 提出之初，“挖掘”的对象是数据库，而目前挖掘的对象已经扩大到文本、网页、流媒体等，因此，数据挖掘的含义已扩大到包括文本挖掘、网页挖掘、流媒体挖掘等在内的范畴。

数据挖掘主要包括预测型和描述型两大类；数据挖掘采用的算法包括决

策树、回归分析、神经网络、关联分析和聚类分析等。在用数据挖掘技术解决问题时，一般先确定该问题是预测型问题还是描述型问题，然后确定采用的算法。

历史数据的变化规律具有一定的准确性、可信性和稳定性。预测型问题解决的方法是从历史数据找出这种变化规律，即建立模型。模型是对历史数据规律的描述，然后用此模型来预测数据的未来发展趋势。所以，通常建立一个好的模型后，可以多次使用对数据的未来进行预测。图 2 为现有技术的预测型方法流程图。如图 2 所示，该方法包括以下步骤：

步骤 210：建模数据处理，手工使用挖掘工具，建立数据挖掘模型。

本步骤包括：搜集与预测问题相关的历史数据，从中选择建模数据，并转换为数据挖掘需要的格式；手工使用挖掘工具，选择适当的挖掘算法，对准备好的建模数据进行处理，获得数据挖掘模型。

步骤 220：评估数据处理。

本步骤包括：搜集与预测问题相关的历史数据，从中选择评估数据，并转换为数据挖掘需要的格式。

步骤 221 ~ 步骤 222：手工使用挖掘工具进行模型评估，得到评估报告，根据评估报告判断模型是否好，好的话执行步骤 231，否则执行步骤 210。

本步骤包括：利用步骤 220 准备好的评估数据，对步骤 210 建立的模型进行评估，即利用建立好的模型对历史数据进行预测，如果预测的结果与历史数据中对应的结果相符合，则认为模型好；如果预测的结果与历史数据中对应的结果相差较大，则认为模型不好。如果模型效果不好，则返回步骤 210 调整建模数据或重新建立模型，如果模型比较好，则可以将模型投入应用。

步骤 230：评分数据处理。

本步骤包括：搜集与预测问题相关的数据，转换为数据挖掘需要的格式。

步骤 231：手工使用挖掘工具，进行评分处理。

本步骤包括：手工使用挖掘工具，利用步骤 211 建立好的模型，对步骤

230 准备好的评分数据进行处理，得到预测结果，即数据的未来发展趋势。

例如，在客户流失模型中，评分结果反映的是客户流失可能性的大小，一般用 0~1 之间的一个数值表示，此值越接近 1 说明客户流失的可能性越大。如本步骤评分处理后得到某些或某个客户的预测结果是 0.8，可以理解为该批客户或该客户的流失可能性为 80%。

步骤 232：手工使用挖掘工具导出预测结果。

本步骤包括：将步骤 231 计算出的预测结果从数据挖掘工具导入到数据库中。

步骤 233：在数据库中对预测结果进行分析，以便对不同特征的数据采取不同的措施。

例如，在客户流失模型中得到某些客户流失的可能性为 80%，即客户流失的可能性比较大，则运营商可以针对该批客户采取一些挽留的措施，以保证该批客户继续为运营商带来利润。

描述型是对一组对象的某些方面或者角度进行特征描述和关联分析等，从中找出一些规律。例如客户细分：由于客户的经济、社会、行为等各种属性都有差别，所以他们的需求各不相同，商家为了吸引更多的客户，需要把客户分成不同的群体，然后对不同的群体采取不同的营销策略，从而在满足客户的同时最大化的获得利润。关联分析，例如在购买面包和黄油为顾客中，有 90% 的人同时也买了牛奶，即面包 + 黄油 -> 牛奶；商家在通过数据挖掘得到此关联结果后，可以将面包同、黄油和牛奶摆放在相邻的位置，以方便客户购买。当需要处理的数据量大时，需要借助挖掘工具建立数据挖掘模型，例如，客户细分时，在目标客户属性多、数据量大时，就需要借助挖掘工具建立数据挖掘的细分模型。进一步，如果细分的客户数据量相当大，数据挖掘需要的计算机系统资源比较多，可能无法建立细分模型时，可以对数据进行抽样，先使用抽样的客户数据建立细分模型，并进行模型评估，得到好的细分模型后，手工使用挖掘工具根据细分模型将其他客户分到相应的分组中，这个过程也称为评分。

图 3 为现有技术描述型方法流程图,如图 3 所示,该方法包括以下步骤:

步骤 310: 建模数据处理。

搜集与预测问题相关的历史数据,从中选择建模数据,并转换为数据挖掘需要的格式。

步骤 311~步骤 312: 判断数据量是否大,如果数据量大,则将数据抽样后,执行步骤 313; 否则直接执行步骤 313。

步骤 313: 手工使用挖掘工具建立数据挖掘模型。

步骤 320~步骤 321: 评估数据处理,手工使用挖掘工具进行模型评估,得到评估报告。

步骤 322: 根据评估报告判断该模型是否好,好的话执行步骤 323, 否则执行步骤 313 或步骤 310。

步骤 323: 判断是否是抽样建模,是则执行步骤 324, 否则执行步骤 326。

步骤 330: 评分数据处理。

本步骤包括: 搜集与描述问题相关的数据,转换为数据挖掘需要的格式。

步骤 324: 手工使用挖掘工具进行模型评分。

步骤 325: 处理得到的评分结果。

步骤 326: 模型发布。

例如,客户细分模型中,得到评分结果后,将细分后的客户名单输出。

目前,挖掘工具使用的是第三方工具,模型建立完成后,模型的应用过程通常是: 在数据库中处理好评分数据,然后使用挖掘工具进行评分操作,再手工使用挖掘工具将评分结果导入数据库,最后在数据库中对评分结果进行分析处理。可见,每次评分过程的实现都要借助于挖掘工具,而一个好的模型建立后,可以稳定使用一段时间,但每次使用模型都无法脱离挖掘工具,需要人力介入,自动化程度不够,且效率比较低,从而影响了数据挖掘的效率,尤其是在模型比较多时更是如此。

现有挖掘工具通常可以将模型以标准 C 文件的形式导出,用开发工具开发一个应用程序,将 C 模型嵌入到应用程序中,由该应用程序实现评分过

程。这种处理方法比较简单，脱离了挖掘工具，提高了模型的使用效率，且有一定的通用性。但是，每次使用新的模型时，要将原来的数据挖掘模型的C源程序替换为新的数据挖掘模型的C源程序，需要重新编译评分应用程序，自动化程序也不够，从而影响了数据挖掘的效率。而且C文件形式的模型一般不容易与基于数据库的数据处理过程融为一体，同时模型的应用依赖于数据挖掘工具本身的设置信息。

发明内容

有鉴于此，本发明实施例提供一种实现数据挖掘模型应用的方法，实现数据挖掘模型的自动化应用，提高数据挖掘的效率。

本发明实施例提供一种实现数据挖掘模型应用的系统，实现数据挖掘模型的自动化应用，提高数据挖掘的效率。

本发明实施例提供一种模型转换装置，实现数据挖掘模型的自动化应用，提高数据挖掘的效率。

为达到上述目的，本发明实施例的技术方案是这样实现的：

一种实现数据挖掘模型应用的方法，该方法包括：导出建立好的数据挖掘模型；

根据预设的数据挖掘模型所使用语言与评分过程使用语言中语句逻辑的对应关系将数据挖掘模型转换为评分过程使用语言的数据挖掘模型；

将转换后的数据挖掘模型保存为存储过程或函数；

将评分数据作为所述存储过程或函数的参数，通过调用所述存储过程或函数来完成评分过程。

所述数据挖掘模型为预测型模型，所述转换包括：

搜索所述数据挖掘模型，当搜索到标识数据挖掘模型结束的字符串时，从所述标识数据挖掘模型结束的字符串所在行逐行向前搜索；

当搜索到转行符时，将转行符连接的两行合并；

当搜索到条件语句时，替换成评分过程使用语言的条件语句；

当搜索到数据挖掘工具根据目标字段生成的字符串时，则将字符串删去，
输出字符串后的常数数值；

搜索到标识数据挖掘模型开始的字符串，则结束流程。

所述数据挖掘模型为预测型模型，所述转换包括：

搜索所述数据挖掘模型，当搜索到标识数据挖掘模型开始的字符串时，从
所述标识数据挖掘模型开始的字符串所在行逐行向后搜索；

当搜索到转行符时，将转行符连接的两行合并；

当搜索到条件语句时，替换成评分过程使用语言的条件语句；

当搜索到数据挖掘工具根据目标字段生成的字符串时，则将字符串删去，
输出字符串后的常数数值；

搜索到标识数据挖掘模型结束的字符串，则结束流程。

所述数据挖掘模型为聚类算法描述型数据挖掘模型，所述转换包括：

搜索所述数据挖掘模型，当搜索到模型中标识细分分组的字符串时，生成
与分组个数相同的分组数组，从模型中标识细分分组的字符串所在行逐行向前
搜索；

搜索到条件语句，利用条件语句中的常数数值计算出评分数据中每个字段
的数据到各个分组中心的距离，输出所述距离到所述分组数组；

搜索到模型中标识细分分组结束的字符串，则结束搜索；

比较每个评分数据到各个分组中心的距离，并将所述评分数据归到距离
最近的分组中。

一种实现数据挖掘模型应用的系统，该系统包括模型建立导出装置、模
型转换装置和评分装置，

所述模型建立导出装置，用于建立数据挖掘模型，将建立好的数据挖掘
模型导出到模型转换装置；

所述模型转换装置，用于保存预先设置的导出的数据挖掘模型所使用的

语言与评分过程使用语言中语句逻辑的对应关系；接收模型建立导出装置导出的数据挖掘模型，根据所述对应关系将所述数据挖掘模型转换为评分过程使用语言的数据挖掘模型，将转换后的数据挖掘模型保存为存储过程或函数，并输出到评分装置；

所述评分装置，用于接收所述模型转换装置输出的所述存储过程或函数，通过调用所述存储过程或函数来完成评分过程。

一种模型转换装置，该装置包括搜索模块、对应关系存储模块和传输模块，

所述传输模块，用于接收模型建立导出装置导出的数据挖掘模型，传输到搜索模块；接收搜索模块传输的转换后的数据挖掘模型，将所述转换后的数据挖掘模型保存为存储过程或函数，并传输到评分装置；

所述搜索模块，用于接收传输模块传输的导出的数据挖掘模型，根据对应关系存储模块中保存的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系，将所述模型建立导出装置导出的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型，将转换后的数据挖掘模型传输到传输模块；

所述对应关系存储模块，用于保存预先设置的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系。

与现有技术相比，本发明实施例所提供的实现数据挖掘模型转换和应用的方法、系统及装置，通过将建立的数据挖掘模型转换为评分过程使用语言的数据挖掘模型，并且自动调用该转换后的数据挖掘模型，实现了数据挖掘模型的自动化应用，提高了数据挖掘的效率。同时，评分过程自动调用转换后的数据挖掘模型，将数据挖掘模型与评分过程的数据处理过程融为一体，脱离了数据挖掘工具中的设置信息。

附图说明

- 图 1 为现有技术 KDD 处理过程；
- 图 2 为现有技术预测型方法流程图；
- 图 3 为现有技术描述型方法流程图；
- 图 4 为本发明实施例中的实现数据挖掘模型应用的方法流程图；
- 图 5 为本发明实施例一预测型数据挖掘模型的建模、评估方法流程图；
- 图 6 为本发明实施例一预测型数据挖掘模型的评分方法流程图；
- 图 7 为本发明实施例一预测型数据挖掘模型中条件语句的转换方法流程图；
- 图 8 为本发明实施例一预测型数据挖掘模型转换的方法流程图；
- 图 9 为本发明实施例二描述型数据挖掘模型的建模、评估方法流程图；
- 图 10 为本发明实施例二描述型数据挖掘模型的评分方法流程图；
- 图 11 为本发明实施例二描述型数据挖掘模型转换的总体方法流程图；
- 图 12 为本发明实施例的实现数据挖掘模型应用的系统结构图。

具体实施方式

下面结合附图及具体实施例对本发明进行详细说明。

本发明实施例的实现数据挖掘模型应用的方法包括：导出建立好的数据挖掘模型；根据预设的数据挖掘模型所使用语言与评分过程使用语言中语句逻辑的对应关系，将数据挖掘模型转换为评分过程所使用语言的数据挖掘模型，并在评分过程中调用该转换后的数据挖掘模型。

图 4 为本发明实施例的实现数据挖掘模型应用的方法流程图。如图 4 所示，该方法包括以下步骤：

步骤 400：将建立好的数据挖掘模型导出；

具体的导出过程与现有技术的处理方式相同，不在本发明关注的范围之内，在此不再赘述。

步骤 401：预先设置数据挖掘工具导出的数据挖掘模型所使用语言与评

分过程所使用语言中语句逻辑的对应关系。

步骤 402: 根据预先设置的数据挖掘工具导出的数据挖掘模型所使用语言与评分过程所使用语言中语言逻辑的对应关系, 将导出的数据挖掘模型转换为评分过程使用语言的数据挖掘模型;

步骤 403: 调用转换后的数据挖掘模型完成评分过程。

该方法还可以进一步包括: 将转换出的评分过程使用语言的数据挖掘模型保存为函数, 或嵌入到预先写好的评分存储过程模板中, 保存为存储过程; 此时, 调用转换后的数据挖掘模型完成评分过程的方法为: 调用该函数或调用该存储过程完成评分过程。

实际应用中, 步骤 401 可在步骤 400 与步骤 402 之间执行, 也可在步骤 400 之前执行。

以下分别就两个实施例对本发明实施例的实现数据挖掘模型应用的方法进行详细说明。

实施例一:

本实施例对预测型数据挖掘模型在数据库中的应用方法进行详细介绍。

以下首先介绍模型转换和利用转换后的模型进行自动评分的过程, 然后介绍模型转换的具体过程。

图 5 为本发明实施一预测型数据挖掘模型建模、评估的方法流程图, 如图 5 所示, 该方法在数据挖掘模型的建立和评估完成后, 包括以下步骤:

步骤 510 ~ 步骤 522 与步骤 210 ~ 步骤 222 相同, 在此不再赘述。

步骤 523: 导出建立好的数据挖掘模型。

具体的导出过程与现有技术的处理方式相同, 不在本发明关注的范围之内, 在此不再赘述。

本步骤将建立好的数据挖掘模型以标准语言的形式导出, 如导出标准的 C 语言形式, 具体导出的数据挖掘模型的语言形式取决于数据挖掘工具本身提供的功能。目前常用的数据挖掘工具都可以将建立好的数据挖掘模型以标准 C 文件的形式导出。

步骤 524: 数据挖掘模型转换。

本步骤包括: 预先设置数据挖掘工具导出的数据挖掘模型使用语言与评分过程使用语言中语句逻辑的对应关系, 根据该对应关系将从挖掘工具导出的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型。

步骤 525: 将转换出的数据挖掘模型保存为存储过程或函数。

在本步骤, 标准语言格式的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型后, 嵌入到预先写好的存储过程模板中, 保存为自动评分存储过程。

本步骤也可将导出的数据挖掘模型保存为函数, 在评分过程中设置调用程序调用该函数完成评分过程。

预先写好的评分存储过程模板是根据评分过程使用的语言的使用规则, 将一个完整软件所需要的框架写好, 预留出填入转换出的数据挖掘模型的空间, 数据挖掘模型转换完成后直接写入预留的空间即可。

在数据库查询语言(SQL)中, 存储过程经编译后驻留在数据库中, 可以被应用程序通过引用其名称而调用, 并允许数据以参数形式在存储过程和应用程序间来回传递。在 SQL 数据库中, 用户可以自定义存储过程。

图 6 为本发明实施例一预测型数据挖掘模型的评分方法流程图, 如图 6 所示, 该方法包括以下步骤:

步骤 600: 评分数据处理。

本步骤执行的具体过程与步骤 230 相同, 在此不再赘述。

步骤 601: 调用自动评分存储过程或函数。

步骤 602: 得到预测结果。

步骤 603: 对结果进行分析。

本步骤是指对预测结果进行分析, 以确定对不同特征的数据采取不同的措施。

评分过程调用自动评分存储过程后, 将评分数据以参数的形式传递到存储过程, 经存储过程计算得到结果, 并将评分结果写入评分过程使用的数据库

库中，供数据库进行分析。

以下以标准 C 语言的统计分析系统(SAS)预测型数据挖掘模型转换为评分过程使用的 SQL 的数据挖掘模型的转换方法为例对本发明实施例进行详细说明。

在本发明实施例中，标准 C 语言的 SAS 预测型数据挖掘模型转换为评分过程使用的 SQL 的数据挖掘模型的具体转换过程包括：

第一，搜索到从数据挖掘工具中导出的数据挖掘模型中标识该数据挖掘模型结束的标志，从标识该数据挖掘模型结束的标志所在行逐行向前搜索。

第二，搜索到转行符，则将转行符连接的两行合并为一行，进行后面的搜索和替换处理。

第三，搜索到条件语句，则将条件语句转换为评分过程使用的语言后输出。

图 7 为本发明实施例一预测型数据挖掘模型中的条件语句转换的方法流程图，如图 7 所示，本步骤将条件语句转换为 SQL 语言中的条件语句的方法包括以下步骤：

步骤 700：从数据挖掘模型中条件语句的最后一行开始向前搜索，如果该行有标识条件语句结束的字符“}”，且该“}”是该条件语句中的第一个，则将第一个“}”用“end”替换。

步骤 701：向前搜索到“else”，则删除“else”后的“{”，保留该“else”。

步骤 702：从“else”继续向前搜索，遇到第一个“}”，则删除，第二个“}”用“end”替换。

步骤 703：继续向前搜索，遇到“if”语句，则用“case when...then”替换，其中“case when...then”的条件是“if”语句后的条件。

第四，搜索到挖掘工具根据目标字段生成的字符串，则将字符串删去，输出字符串后的常数数值。

例如，如果在 SAS 客户流失数据挖掘模型中搜索到根据目标字段“CHURN”生成的字符串“P_CHURN1=0.025559”，则删除该字符串，保

留该字符串后面的常数数值，即输出常数数值 0.025559。

第五，搜索到数据挖掘模型中标识该数据挖掘模型结束的字符串，则结束数据挖掘模型转换流程。

客户流失分析是数据挖掘中预测问题的一种具体应用，目前采用的数据挖掘算法常用的有决策树、线性回归等算法。以下以 SAS 决策树数据挖掘模型转换为评分过程使用的 SQL 的数据挖掘模型的具体转换过程进行详细说明。

实际应用中，数据挖掘模型转换时也可以通过从前往后搜索的方法实现。图 8 为本发明实施例一预测型数据挖掘模型转换的方法流程图。该图中，数据挖掘模型转换过程是通过从前往后搜索而实现的。如图 8 所示，该方法包括以下步骤：

步骤 800 ~ 步骤 802: 逐行读取 SAS 决策树数据挖掘模型，判断该行是否是标识数据挖掘模型开始的字符串，不是则读取下一行，是则执行步骤 803。

步骤 803: 读标识数据挖掘模型开始的字符串的下一行。

步骤 804 ~ 步骤 805: 如果有换行符，则将换行符连接的两行合并为一行后，执行步骤 806; 如果没有换行符，直接执行步骤 806。

步骤 806 ~ 步骤 807: 判断该行是否是标识数据挖掘模型结束的字符串，如果是，则结束流程; 否则执行步骤 808。

步骤 808: 判断该行是否是“if”语句，如果是，则执行步骤 809; 否则执行步骤 810。

步骤 809: 输出“case when ... then”，然后执行步骤 803，读下一行。

步骤 810: 判断该行是否有“{”，如果有“{”，则执行步骤 811; 如果没有“{”，则执行步骤 815。

步骤 811: 判断该行是否有“else”，如果有“else”，则执行步骤 812; 如果没有“else”则执行步骤 813。

步骤 812: 将“else”推进栈，并输出，然后执行步骤 803，读下一行。

步骤 813: 判断前一行是否是“if”语句, 如果是“if”语句, 则执行步骤 814; 如果不是“if”语句, 则执行步骤 803, 读下一行。

步骤 814: 将“if”推进栈, 然后执行步骤 803, 读下一行。

步骤 815: 判断该行是否有 SAS 决策树数据挖掘工具根据目标字段生成的字符串, 如果有, 则执行步骤 816; 如果没有, 则执行步骤 817。

步骤 816: 输出该字符串后的数值, 然后执行步骤 803, 读下一行。

例如, 如果是字符串“P_CHURN1=0.025559”, 则输出字符串后的数值 0.025559。

步骤 817: 判断该行是否有“}”, 如果有“}”, 则执行步骤 818; 如果没有则执行步骤 803, 读下一行

步骤 818: 推出栈顶的内容。

步骤 819: 判断步骤 818 中出栈的内容是否是“else”, 如果是则执行步骤 820; 否则执行步骤 803, 读下一行。

步骤 820: 输出“end”, 然后执行步骤 803, 读下一行。

实际应用中, 该流程可以由任何一种编程语言实现。

实施例二:

本实施例对描述型数据挖掘模型在数据库中应用的方法进行详细介绍。

以下首先介绍数据挖掘模型转换和利用转换后的数据挖掘模型进行评分的过程, 然后介绍数据挖掘模型转换的具体过程。

图 9 所示为本发明实施例二描述型数据挖掘模型的建模、评估方法流程图。如图 9 所示, 该方法包括以下步骤:

步骤 910~步骤 922 与步骤 310~步骤 322 相同, 在此不再赘述。

步骤 923: 确定是抽样建模。

例如, 客户细分模型, 客户数据量大, 建模时需要进行抽样, 根据抽样数据建立数据挖掘模型, 然后利用建立好的数据挖掘模型再进行其他客户数据的细分。所以, 在本步骤, 需要确定是否是抽样建模, 如果是则执行以下流程, 否则结束流程。

步骤 924: 导出建立好的数据挖掘模型。

将建立好的数据挖掘模型以标准语言的形式输出, 如以标准 C 语言的形式输出。

具体的导出过程与现有技术的处理方式相同, 不在本发明关注的范围之内, 在此不再赘述。

步骤 925: 数据挖掘模型转换。

根据预先设置的数据挖掘工具使用语言与评分过程使用语言中语句逻辑的对应关系将导出的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型。

步骤 926: 将转换出的数据挖掘模型保存为自动评分存储过程或函数。

本步骤中, 将转换出的数据挖掘模型嵌入到预先写好的自动评分存储过程模板中, 保存为自动评分存储过程, 以便于评分时调用该存储过程, 实现自动评分过程。

图 10 为本发明实施例二描述型数据挖掘模型的评分方法流程图。如图 10 所示, 该方法包括以下步骤:

步骤 100: 评分数据处理。

本步骤的实现方法与步骤 330 相同, 在此不再赘述。

步骤 1001: 调用自动评分存储过程。

步骤 1002: 数据挖掘模型发布。

例如, 在客户细分模型中, 评分过程完成后得到所有客户分类的结果, 商家可对不同类型的客户采取不同的营销策略。

聚类(Cluster)是目前客户细分模型中常用的方法。本发明实施例的标准 C 语言的描述型聚类数据挖掘模型转换为评分过程使用的数据库语言 SQL 的具体转换过程包括:

第一, 搜索到模型中标识细分分组的字符串, 生成与分组个数相同的分组数组, 如 `cluster[j]`, 其中 `j` 为分组个数; 并从该行向前搜索。

第二, 搜索到条件语句, 利用条件语句中的常数数值计算出评分数据中

每个字段到各个分组中心的距离，然后输出该距离到建立的分组数组中。

本步骤，把条件语句中的每一个字段按照这样的方式累加：

$$\text{Cluster}[j] = \sum_{ji} (Fi - Bi)^2$$

其中， F_i 是第*i*个字段， B_i 是对应分组中的条件语句下面的常数，意思是分组在该字段上的中心。然后输出，得到的结果即为评分数据中每个字段到各个分组中心的距离。

如果在数据挖掘工具算法中有是否进行标准化处理的选项，而在建立数据挖掘模型中选择了该选项，则利用条件语句后面第一个语句中的常数和各个分组在各字段上的中心值计算出评分数据中每个字段到各个分组中心的距离，输出该距。即，把条件语句中每一个字段的累加格式为：

$$\text{Cluster}[j] = \sum_{ji} ((Fi * Ai) - Bi)^2$$

其中， A_i 是条件语句后面第一个语句中的常数，是标准化处理系数。

第三，搜索到 SAS 工具根据离散属性变量生成的字符串，则搜索到该离散属性的变量，将字符串替换为该变量。

如果 SAS 模型中条件语句后面的字段是 D_i 的形式，这是 SAS 工具处理时对离散属性替换的变量，这时要搜索替换的原字段，可以通过搜索与 D_i 对应的条件语句获得。

例如，离散变量 X 有三个离散值，分别为 2, 5, 6，则 SAS 工具处理该离散属性的变量时，会针对这个离散变量 X 生成 D_1，D_1 有三个变量值分别为 2, 5, 6。如果要搜索替换的原字段 X，则要搜索 D_1 对应的条件语句得到 X，并用 X 替换 D_1。

第四，搜索到模型中标识细分分组结束的字符串，则搜索结束。

如搜索到 “SEGMENT=1” 则搜索结束。

这样就获得了各个分组评分的距离标准，可以对每个抽样数据进行处理，也就是计算每个数据与分组中心的距离。

第五，比较每个评分数据距离各个分组中心的距离，并归到距离最近的

分组中。

以下以标准 C 语言的聚类数据挖掘模型转换为评分过程使用的数据库语言 SQL 的自动转换方法为例做进一步详细说明。

图 11 为本发明实施例二描述型数据挖掘模型转换的总体方法流程图。如图 11 所示，该方法包括以下步骤：

步骤 1100 ~ 步骤 1102：开始，逐行读取 SAS 聚类数据挖掘模型，判断该行是否是标识细分分组开始的字符串，不是则读取下一行，是则执行步骤 1103。

步骤 1103：模型开始后，建立与分组个数相同的分组数组。

步骤 1104：读数据挖掘模型的下一行。

步骤 1105 ~ 步骤 1106：如果该行有换行符，则合并换行符连接的两行后，执行步骤 1107；否则直接执行步骤 1107。

步骤 1107：判断该行是否是标识细分分组结束的字符串，如果是，则执行步骤 1108；否则执行步骤 1110。

步骤 1108：将评分数据各字段的数据到各个分组中心的距离输出到建立的分组数组中。

步骤 1109：比较距离，输出分组。

步骤 1110：判断该行是否是 if 语句，如果是，则执行步骤 1111；否则执行步骤 1104。

步骤 1111：累加分组的距离，然后执行步骤 1107。

以下对实现本发明实施例所提供的实现数据挖掘模型应用的装置进行详细介绍。

图 12 为本发明实施例的实现数据挖掘模型应用的系统结构图。如图 12 所示，该系统包括模型建立导出装置、模型转换装置和评分装置。

其中，模型建立导出装置，用于建立数据挖掘模型，将建立好的数据挖掘模型导出到模型转换装置。

模型转换装置，用于预先设置导出的数据挖掘模型所使用的语言与评分

过程使用语言中语句逻辑的对应关系；接收模型建立导出装置导出的数据挖掘模型，根据所述对应关系，将所述导出的数据挖掘模型转换为评分装置使用语言的数据挖掘模型，将所述转换后的数据挖掘模型输出到评分装置。

评分装置，用于接收所述模型转换装置转换后的数据挖掘模型，调用所述数据挖掘模型完成评分过程。

模型转换装置包括搜索模块、对应关系存储模块和传输模块。

其中，传输模块，用于接收模型建立导出装置导出的数据挖掘模型，传输到搜索模块；接收搜索模块传输的转换后的数据挖掘模型，传输到评分装置；

搜索模块，用于接收传输模块传输的导出的数据挖掘模型，根据对应关系存储模块中保存的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系，将模型建立导出装置导出的数据挖掘模型转换为评分过程所使用语言的数据挖掘模型，将转换后的数据挖掘模型传输到传输模块；

对应关系存储模块，用于保存预先设置的导出的数据挖掘模型所使用的语言与评分过程使用语言中语句逻辑的对应关系。

传输模块进一步用于将所述转换后的数据挖掘模型保存为自动评分存储过程或函数。

评分装置包括评分模块和调用模块。

其中，调用模块，用于接收并保存模型转换装置传输的转换后的数据挖掘模型，将转换后的数据挖掘模型传输到评分模型；

评分模块，用于接收调用模块传输的转换后的数据挖掘模型，完成评分过程。

实际应用中，所有的数据挖掘工具都可以将其建立的模型以标准语言的形式导出。只要是标准语言，就会有一定的规律可循，评分过程使用的语言也是有规律可循的，将数据挖掘工具导出的数据挖掘模型使用语言与评分过程使用语言的对应关系找出后，就可以将挖掘工具导出的数据挖掘模型转换

为评分过程使用语言的数据挖掘模型，进而调用该转换后的数据挖掘模型，实现数据挖掘模型的自动化应用。其转换过程与本发明实施例中所述的自动转换的原理完全相同。

由以上所述可以看出，本发明实施例所提供的一种实现数据挖掘模型转换和应用的方法、系统及装置，将数据挖掘工具导出的数据挖掘模型转换为评分过程使用语言的数据挖掘模型，并且在评分过程中自动调用该转换出的数据挖掘模型，实现了评分过程中数据挖掘模型的自动调用，不需人工干预，从而实现数据挖掘模型的自动化应用。此后，评分处理过程可由其他工具自动调用，无需专门的人员跟踪处理，大大提高了数据挖掘的效率，节省了人力。而且，利用评分过程自动调用存储过程或函数，有利于将数据挖掘模型与评分过程的数据处理过程融为一体，同时，也脱离了数据挖掘工具本身的设置信息，使得数据挖掘能够脱离数据挖掘工具得到应用，提高了数据挖掘的效率。

以上所述仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

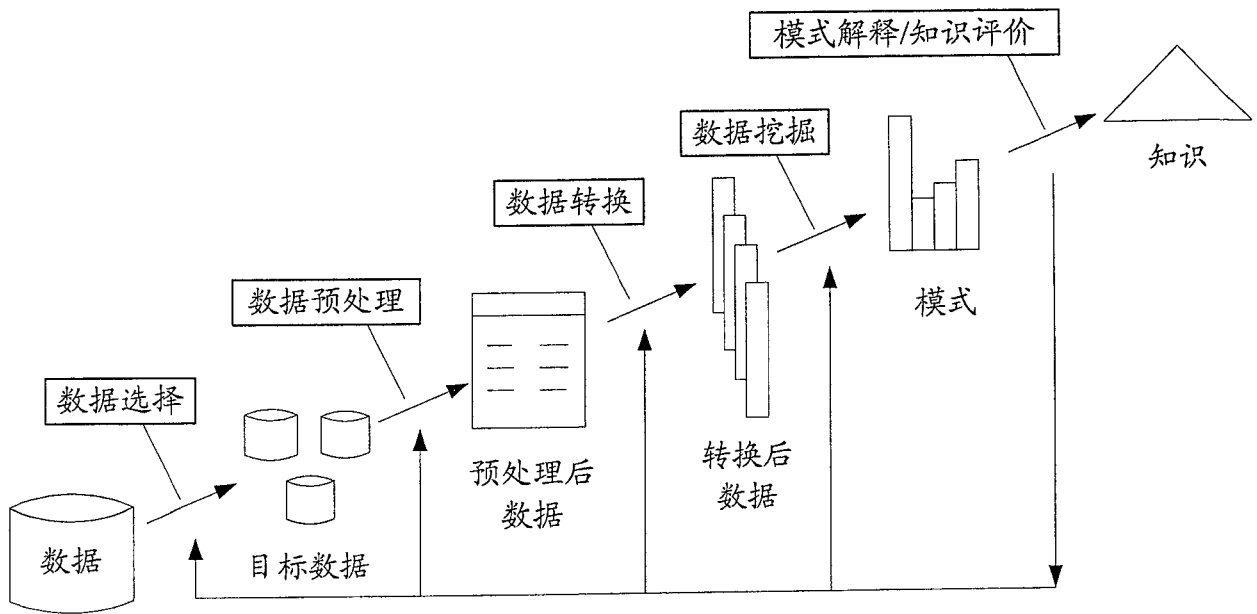


图 1

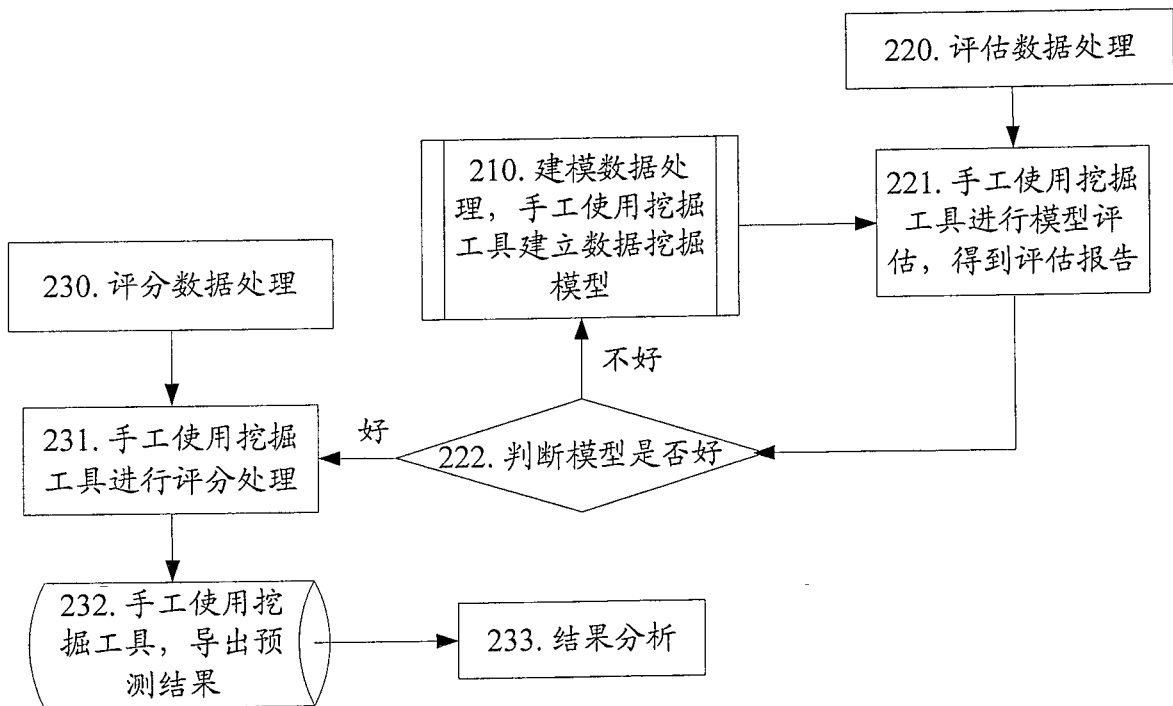


图 2

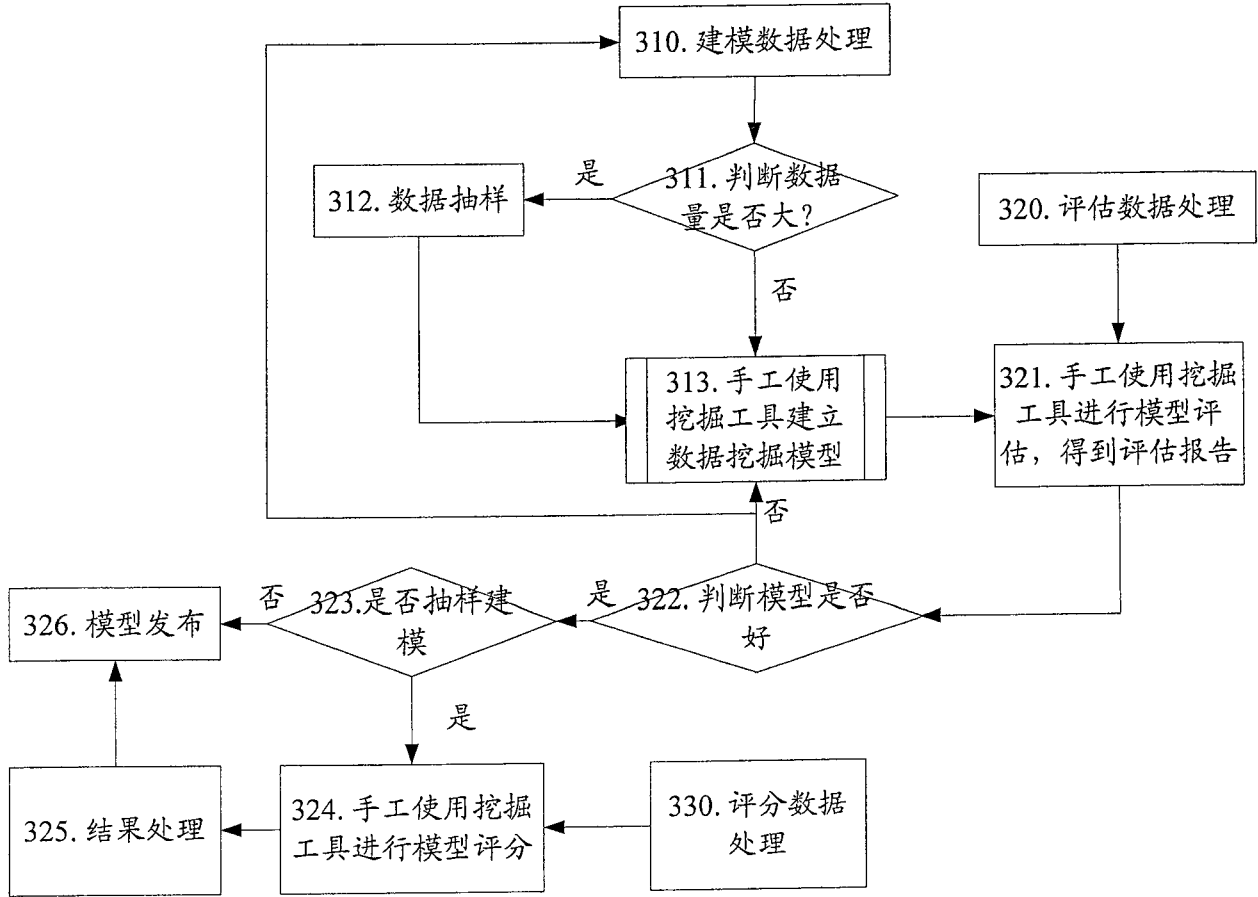


图 3

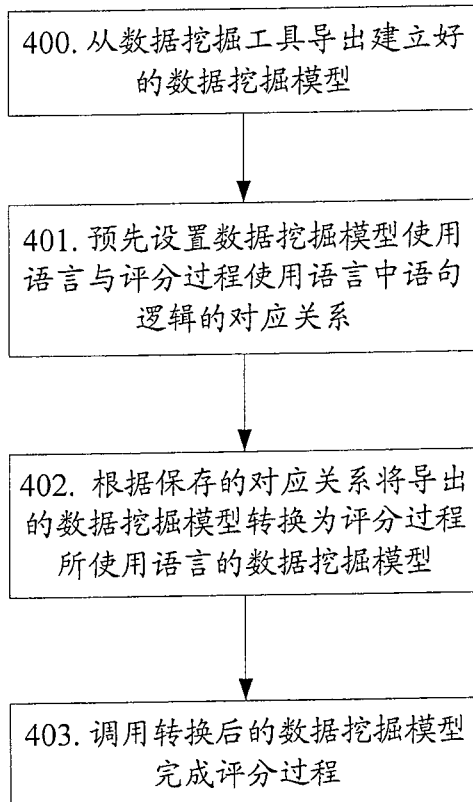


图 4

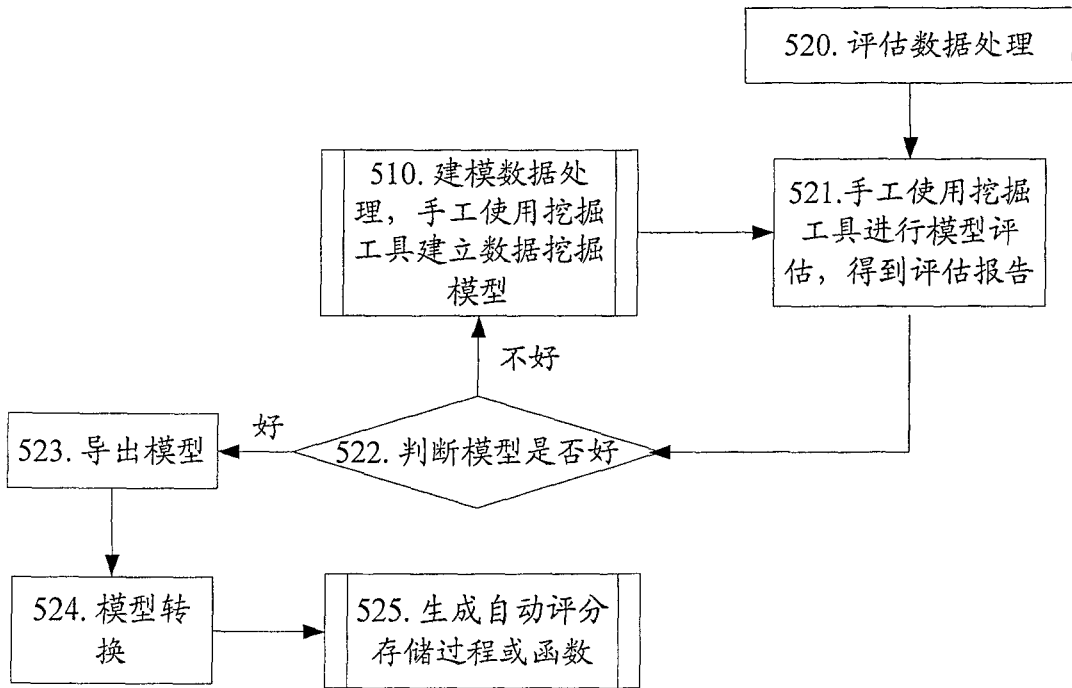


图 5

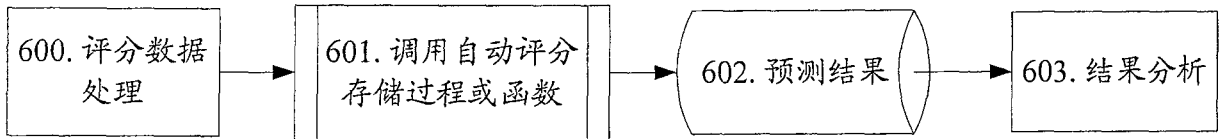


图 6

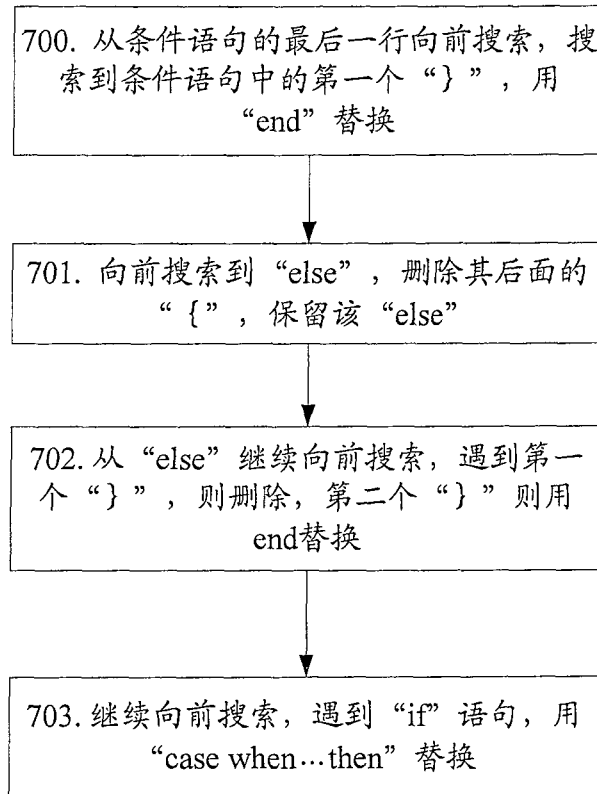


图 7

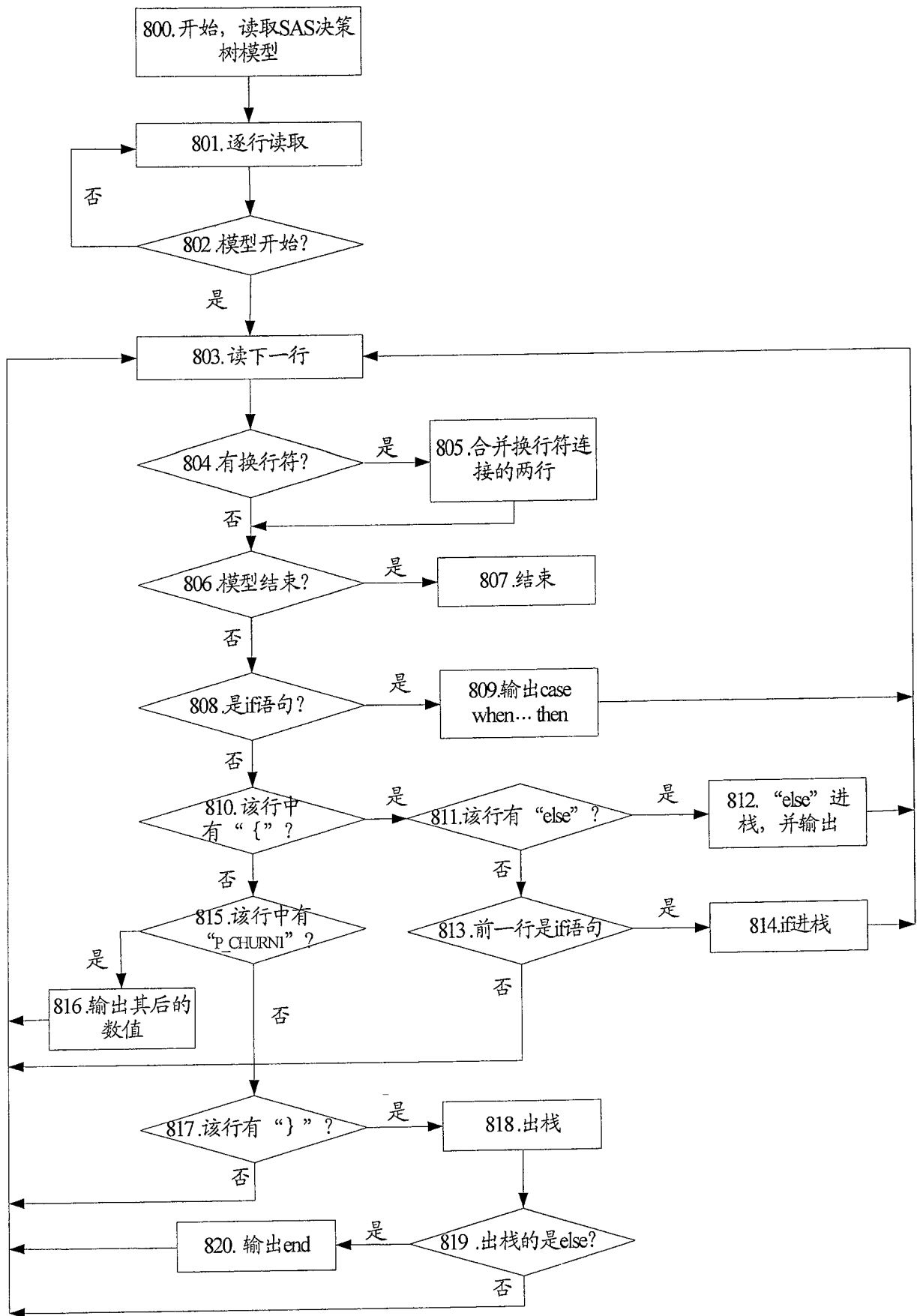


图 8

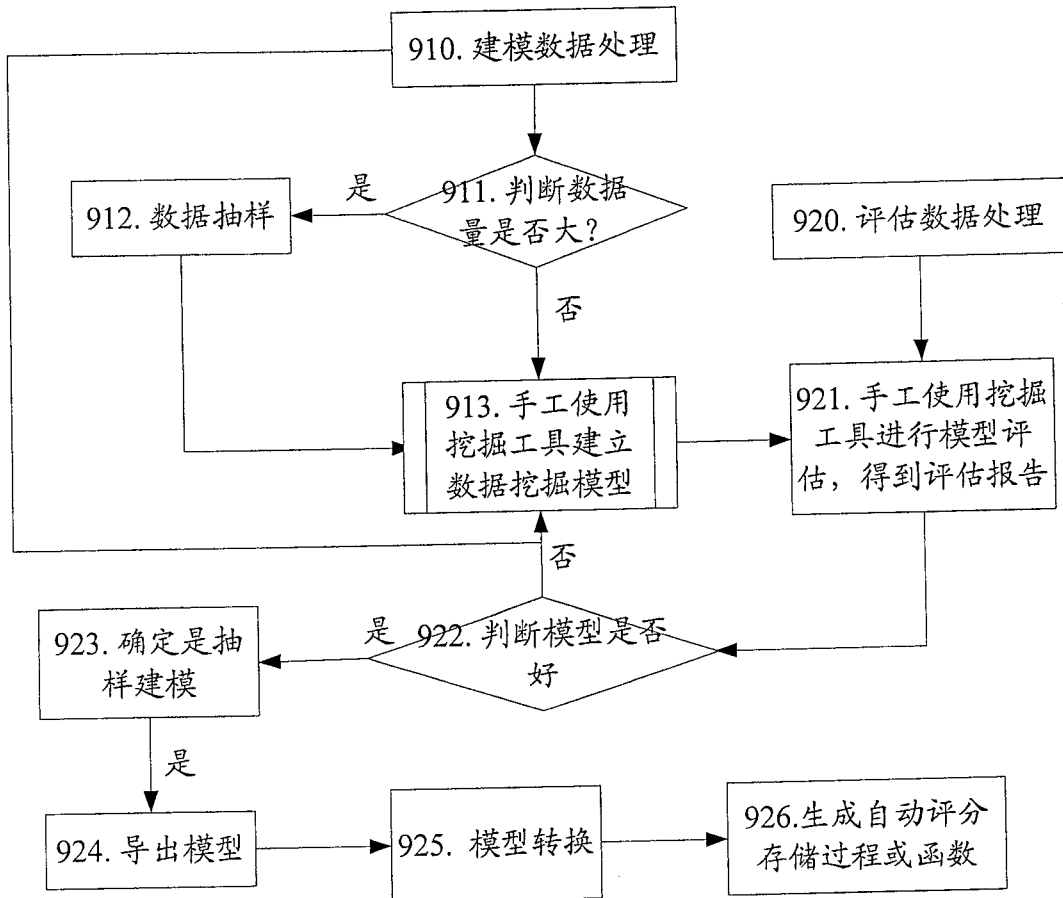


图 9

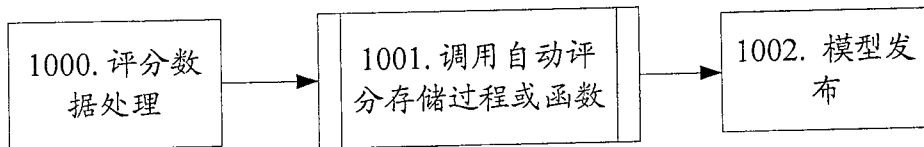


图 10

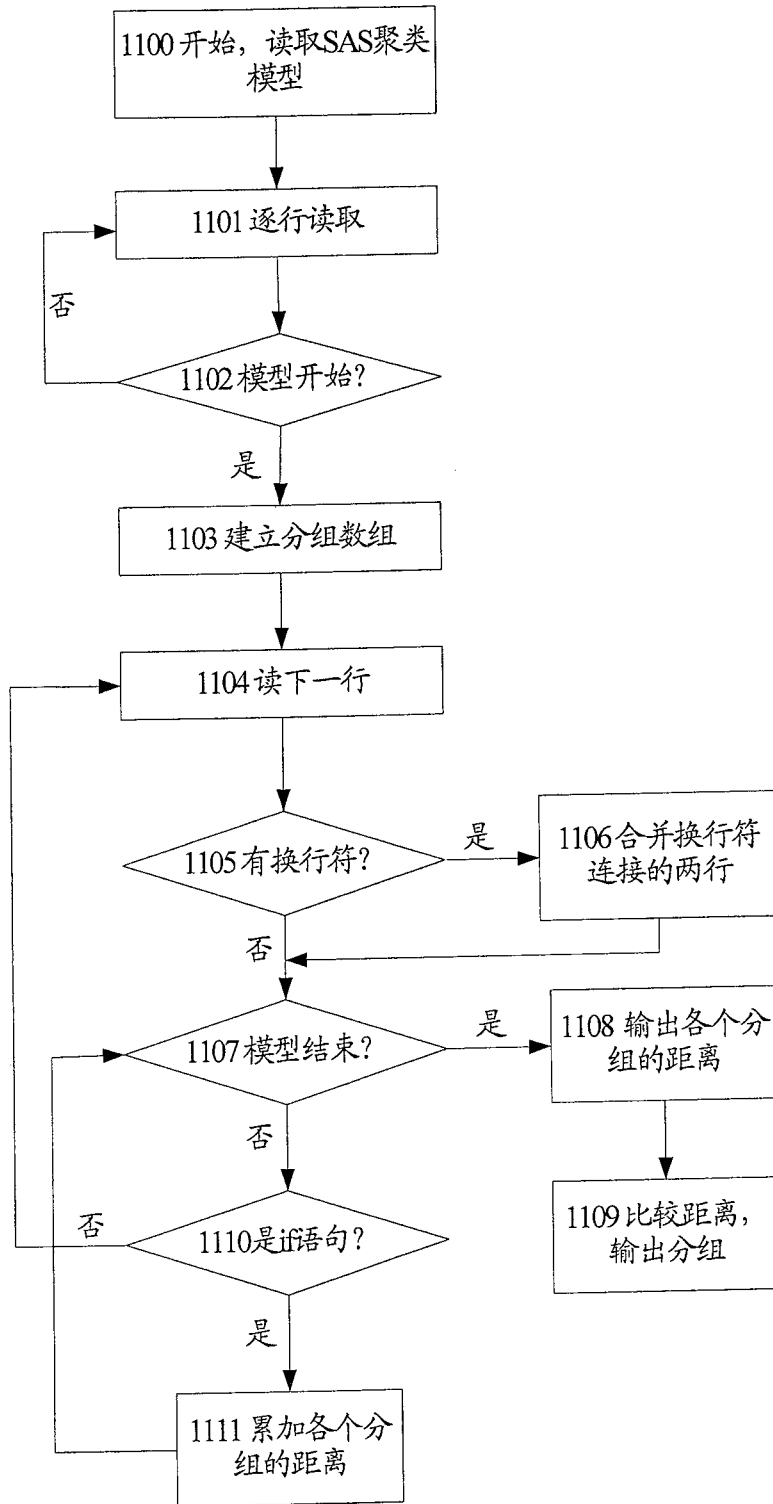


图 11

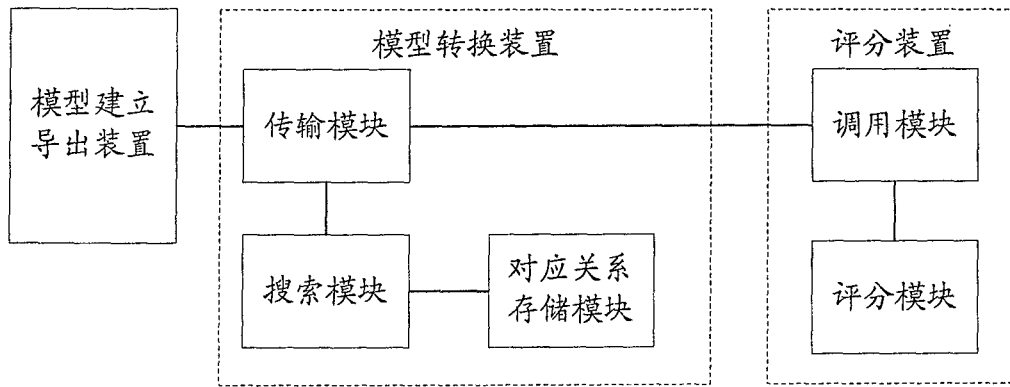


图 12