



# (12)发明专利申请

(10)申请公布号 CN 110990053 A

(43)申请公布日 2020.04.10

(21)申请号 201911225347.5

(22)申请日 2019.12.04

(71)申请人 第四范式(北京)技术有限公司

地址 100085 北京市海淀区上地东路35号

颐泉汇大厦写字楼A座610室

(72)发明人 孔维 宋尧 王萌 吕白芸

朱晓丹 李冠琳 黄纓宁 周振华

(74)专利代理机构 北京展翼知识产权代理事务

所(特殊普通合伙) 11452

代理人 王明远

(51)Int.Cl.

G06F 8/71(2018.01)

G06N 20/00(2019.01)

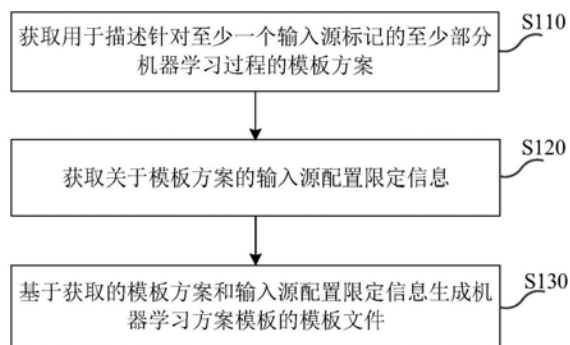
权利要求书2页 说明书23页 附图5页

## (54)发明名称

机器学习方案模板的创建方法、使用方法及装置

## (57)摘要

公开了一种机器学习方案模板的创建方法、使用方法及装置。获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案；获取关于模板方案的输入源配置限定信息，其中，输入源配置限定信息用于生成输入源配置界面，使得经由输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记；以及基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。由此，可以通过复用所创建的机器学习方案模板降低建模门槛以及建模耗时，并且可以解决实际业务数据与模板方案中固有数据之间的数据匹配问题，使得机器学习方案模板在应用于同一业务方向下不同数据结构的业务数据时，均能够获取较好的建模效果。



1. 一种用于创建机器学习方案模板的方法,包括:

获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案,其中,机器学习过程涉及模型训练和/或模型应用;

获取关于所述模板方案的输入源配置限定信息,其中,所述输入源配置限定信息用于生成输入源配置界面,使得经由所述输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记;以及

基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

2. 根据权利要求1所述的方法,其中,获取关于所述模板方案的输入源配置限定信息的步骤包括:

基于获取的模板方案来产生用于设置输入源配置限定信息的控件;

向第一用户展示产生的控件;以及

接收第一用户通过所述控件所设置的输入源配置限定信息。

3. 根据权利要求2所述的方法,其中,

所述输入源标记用于标识模板方案中能够被替换的输入表和/或字段,所述输入源配置限定信息包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各字段的名称、各字段在输入源配置界面上是否展示为可选字段的指示信息。

4. 根据权利要求3所述的方法,其中,

所述输入源配置限定信息还包括各字段对应的字段格式,其中,至少一个字段对应的字段格式被设置为允许针对单个字段配置实际业务数据中的一个或多个字段,使得所配置的一个或多个字段均按照模板方案中处理所述单个字段的同样方式进行字段处理。

5. 根据权利要求3所述的方法,其中,

所述输入源配置限定信息还包括用于限定经由所述输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项。

6. 一种基于机器学习方案模板执行机器学习过程的方法,包括:

获取机器学习方案模板的模板文件,其中,所述模板文件包括模板方案和输入源配置限定信息,所述模板方案用于描述针对至少一个输入源标记的至少部分机器学习过程,所述机器学习过程涉及模型训练和/或模型应用,所述输入源配置限定信息用于生成输入源配置界面;

向第二用户展示基于输入源配置限定信息而生成的输入源配置界面;

获取第二用户经由所述输入源配置界面而配置的至少一个配置输入源;

用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案;

基于修改后的模板方案来执行机器学习过程。

7. 一种用于创建机器学习方案模板的装置,包括:

第一获取模块,用于获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案,其中,机器学习过程涉及模型训练和/或模型应用;

第二获取模块,用于获取关于所述模板方案的输入源配置限定信息,其中,所述输入源配置限定信息用于生成输入源配置界面,使得经由所述输入源配置界面而配置的至少一个

配置输入源替换模板方案中的至少一个输入源标记;以及

生成模块,用于基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

8.一种基于机器学习方案模板执行机器学习过程的装置,包括:

第一获取模块,用于获取机器学习方案模板的模板文件,其中,所述模板文件包括模板方案和输入源配置限定信息,所述模板方案用于描述针对至少一个输入源标记的至少部分机器学习过程,所述机器学习过程涉及模型训练和/或模型应用,所述输入源配置限定信息用于生成输入源配置界面;

第一展示模块,用于向第二用户展示基于输入源配置限定信息而生成的输入源配置界面;

第二获取模块,用于获取第二用户经由所述输入源配置界面而配置的至少一个配置输入源;

第一替换模块,用于用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案;

执行模块,用于基于修改后的模板方案来执行机器学习过程。

9.一种包括至少一个计算装置和至少一个存储指令的存储装置的系统,其中,所述指令在被所述至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求1到6中的任一权利要求所述的方法。

10.一种存储指令的计算机可读存储介质,其中,当所述指令被至少一个计算装置运行时,促使所述至少一个计算装置执行如权利要求1到6中的任一权利要求所述的方法。

## 机器学习方案模板的创建方法、使用方法及装置

### 技术领域

[0001] 本发明总体说来涉及人工智能领域,更具体地说,涉及一种机器学习方案模板的创建方法、使用方法及装置。

### 背景技术

[0002] 机器学习是人工智能研究发展到一定阶段的必然产物,其致力于通过计算的手段,利用经验来改善系统自身的性能。在计算机系统中,“经验”通常以“数据”形式存在,通过机器学习算法,可从数据中产生“模型”,也就是说,将经验数据提供给机器学习算法,就能基于这些经验数据产生模型,在面对新的情况时,模型会提供相应的判断,即,预测结果。可以看出,如何基于经验数据产生模型(即,机器学习建模过程)是机器学习技术的关键。

[0003] 目前在将机器学习技术应用到具体业务场景中时,通常需要建模人员根据业务场景特点,从零开始进行建模调研的工作。从数据准备到模型调试的过程需要较多的时间耗费且对建模人员的素养要求较高,对于业务而言不仅费时,而且有较高的门槛要求。

### 发明内容

[0004] 本发明的示例性实施例旨在克服现有技术中机器学习建模过程耗时且门槛要求较高的缺陷。

[0005] 根据本发明的第一个方面,提出了一种用于创建机器学习方案模板的方法,包括:获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案,其中,机器学习过程涉及模型训练和/或模型应用;获取关于所述模板方案的输入源配置限定信息,其中,所述输入源配置限定信息用于生成输入源配置界面,使得经由所述输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记;以及基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

[0006] 可选地,获取关于所述模板方案的输入源配置限定信息的步骤包括:基于获取的模板方案来产生用于设置输入源配置限定信息的控件;向第一用户展示产生的控件;以及接收第一用户通过所述控件所设置的输入源配置限定信息。

[0007] 可选地,所述输入源标记用于标识模板方案中能够被替换的输入表和/或字段,所述输入源配置限定信息包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各字段的名称、各字段在输入源配置界面上是否展示为可选字段的指示信息。

[0008] 可选地,所述输入源配置限定信息还包括各字段对应的字段格式,其中,至少一个字段对应的字段格式被设置为允许针对单个字段配置实际业务数据中的一个或多个字段,使得所配置的一个或多个字段均按照模板方案中处理所述单个字段的同样方式进行字段处理。

[0009] 可选地,所述输入源配置限定信息还包括用于限定经由所述输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的

处理项。

[0010] 可选地,所述处理项包括关于各字段的校验项,其中,校验项包括各字段的允许格式和/或允许取值范围。

[0011] 可选地,校验项还包括是否进行校验的指示信息。

[0012] 可选地,所述模板方案包括至少一个参数占位符,并且,所述方法还包括:获取关于模板方案的参数配置限定信息,其中,所述参数配置限定信息用于生成参数配置界面,使得经由所述参数配置界面而配置的至少一个配置参数替换模板方案中的至少一个参数占位符,所述基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件的步骤包括:基于获取的模板方案、参数配置限定信息和输入源配置限定信息来生成机器学习方案模板的模板文件。

[0013] 可选地,所述参数配置限定信息包括以下项之中的至少一个:参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值。

[0014] 可选地,所述类型信息指示脚本参数和/或运行参数。

[0015] 可选地,所述参数配置限定信息还包括用于限定经由所述参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项。

[0016] 可选地,所述处理项包括对所述至少一个配置参数进行校验的校验项。

[0017] 可选地,所述获取关于模板方案的参数配置限定信息的步骤包括:基于获取的模板方案来产生用于设置参数配置限定信息的控件;向第一用户展示产生的控件;以及接收第一用户通过所述控件所设置的参数配置限定信息。

[0018] 可选地,所述参数配置限定信息还包括:用于限定在参数配置界面上按照分类区域对参数占位符进行配置的分类信息,其中,所述获取关于模板方案的参数配置限定信息的步骤还包括:向第一用户展示用于对参数占位符进行分类的控件;以及根据第一用户通过所述控件对参数占位符进行的分类来获取分类信息。

[0019] 可选地,该方法还包括:向第一用户展示用于上传说明文档的控件;接收第一用户通过所述控件所上传的说明文档;以及将说明文档合并入模板文件。

[0020] 可选地,该方法还包括:向第一用户展示用于设置资源配置信息的控件;接收第一用户通过所述控件所设置的资源配置信息,所述资源配置信息用于表征执行所述至少部分机器学习过程的资源配置;以及将资源配置合并入模板文件。

[0021] 可选地,该方法还包括:向第三用户展示基于输入源配置限定信息生成的输入源配置界面;获取第三用户基于与测试场景对应的测试数据表,经由输入源配置界面所配置的至少一个配置输入源;用所述至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的机器学习方案模板;基于修改后的机器学习方案模板执行所述至少部分机器学习过程,以得到所述至少部分机器学习过程的执行结果;对所述执行结果进行评估,以得到测试结果;基于所述测试结果确定是否发布所述机器学习方案模板,或者基于所述测试结果对所述机器学习方案模板进行调试。

[0022] 根据本发明的第二个方面,提出了一种基于机器学习方案模板执行机器学习过程的方法,包括:获取机器学习方案模板的模板文件,其中,所述模板文件包括模板方案和输入源配置限定信息,所述模板方案用于描述针对至少一个输入源标记的至少部分机器学习

过程,所述机器学习过程涉及模型训练和/或模型应用,所述输入源配置限定信息用于生成输入源配置界面;向第二用户展示基于输入源配置限定信息而生成的输入源配置界面;获取第二用户经由所述输入源配置界面而配置的至少一个配置输入源;用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案;基于修改后的模板方案来执行机器学习过程。

[0023] 可选地,所述输入源配置界面中包括用于设置配置输入源的控件,获取第二用户经由所述输入源配置界面而配置的至少一个配置输入源的步骤包括:接收第二用户通过所述控件所设置的配置输入源。

[0024] 可选地,所述输入源标记用于标识模板方案中能够被替换的输入表和/或字段,所述输入源配置限定信息包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各第一字段的名称、各第一字段在输入源配置界面上是否展示为可选字段的指示信息,所述输入源配置界面中还包括以下项之中的至少一个:需要配置的至少一个输入表名称、各输入表下需要配置的各第一字段的名称、各第一字段是否为可选字段的指示信息,所述配置输入源包括以下项之中的至少一个:针对所述输入表配置的业务数据表、针对所述第一字段配置的业务数据表下的第二字段。

[0025] 可选地,所述输入源配置限定信息还包括各第一字段对应的字段格式,其中,至少一个第一字段对应的字段格式被设置为允许针对单个第一字段配置实际业务数据中的一个或多个第二字段,使得所配置的一个或多个第二字段均按照模板方案中处理所述单个第一字段的同样方式进行字段处理,所述输入源配置界面中用于设置针对所述第一字段的第二字段的控件是基于所述第一字段对应的字段格式生成的,以使得第二用户通过该控件能够按照第一字段对应的字段格式针对第一字段配置第二字段。

[0026] 可选地,所述输入源配置限定信息还包括用于限定经由所述输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项,该方法还包括:按照所述处理项对所述配置输入源进行处理;以及在所述输入源配置界面中展示处理结果。

[0027] 可选地,所述处理项包括关于各第一字段的校验项,其中,校验项包括各第一字段的允许格式和/或允许取值范围,按照所述处理项对所述配置输入源进行处理的步骤包括:按照所述第一字段的校验项,对所述第一字段配置的第二字段的格式和/或取值进行校验,其中,所述处理结果用于指示针对所述第一字段配置的第二字段的格式和/或取值是否符合所述校验项。

[0028] 可选地,校验项还包括是否进行校验的指示信息。

[0029] 可选地,所述模板方案包括至少一个参数占位符,所述模板文件还包括参数配置限定信息,所述参数配置限定信息用于生成参数配置界面,该方法还包括:向第二用户展示基于参数配置限定信息而生成的参数配置界面;获取第二用户经由所述参数配置界面而配置的至少一个配置参数;用获取的至少一个配置参数替换模板方案中的至少一个参数占位符,以得到修改后的模板方案。

[0030] 可选地,所述参数配置界面中包括用于设置配置参数的控件,获取第二用户经由所述参数配置界面而配置的至少一个配置参数的步骤包括:接收第二用户通过所述控件所

设置的配置参数。

[0031] 可选地,所述参数配置限定信息包括以下项之中的至少一个:参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值,所述参数配置界面中还包括以下项之中的至少一个:需要配置的参数占位符的类型信息、输入方式信息、展示信息、默认取值。

[0032] 可选地,所述参数配置限定信息还包括用于限定经由所述参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项,该方法还包括:按照所述处理项对所述配置参数进行处理;以及在所述输入源配置界面中展示处理结果。

[0033] 可选地,所述处理项包括对所述至少一个配置参数进行校验的校验项。

[0034] 可选地,所述类型信息指示脚本参数和/或运行参数。

[0035] 可选地,所述参数配置限定信息还包括:用于限定在参数配置界面上按照分类区域对参数占位符进行配置的分类信息,所述参数配置界面按照所述分类信息将需要配置的参数占位符进行分类显示,不同分类的参数占位符被显示在不同分类区域。

[0036] 可选地,所述模板文件还包括用于辅助第二用户了解和/或配置所述模板方案的说明文档,该方法还包括:向所述第二用户提供所述说明文档。

[0037] 可选地,所述模板文件还包括资源配置信息,所述资源配置信息用于表征执行所述至少部分机器学习过程的资源配置,基于修改后的模板方案来执行机器学习过程的步骤包括:基于修改后的模板方案,使用所述资源配置信息所表征的资源配置执行所述机器学习过程,或者基于修改后的模板方案来执行机器学习过程的步骤包括:对修改后的模板方案在执行机器学习过程中所需的资源配置进行预测,使用预测得到的资源配置执行所述机器学习过程。

[0038] 根据本发明的第三个方面,提出了一种用于创建机器学习方案模板的装置,包括:第一获取模块,用于获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案,其中,机器学习过程涉及模型训练和/或模型应用;第二获取模块,用于获取关于所述模板方案的输入源配置限定信息,其中,所述输入源配置限定信息用于生成输入源配置界面,使得经由所述输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记;以及生成模块,用于基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

[0039] 可选地,所述第二获取模块:基于获取的模板方案来产生用于设置输入源配置限定信息的控件;向第一用户展示产生的控件;以及接收第一用户通过所述控件所设置的输入源配置限定信息。

[0040] 可选地,所述输入源标记用于标识模板方案中能够被替换的输入表和/或字段,所述输入源配置限定信息包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各字段的名称、各字段在输入源配置界面上是否展示为可选字段的指示信息。

[0041] 可选地,所述输入源配置限定信息还包括各字段对应的字段格式,其中,至少一个字段对应的字段格式被设置为允许针对单个字段配置实际业务数据中的一个或多个字段,使得所配置的一个或多个字段均按照模板方案中处理所述单个字段的同样方式进行字段

处理。

[0042] 可选地,所述输入源配置限定信息还包括用于限定经由所述输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项。

[0043] 可选地,所述处理项包括关于各字段的校验项,其中,校验项包括各字段的允许格式和/或允许取值范围。

[0044] 可选地,校验项还包括是否进行校验的指示信息。

[0045] 可选地,所述模板方案包括至少一个参数占位符,并且,该装置还包括:第三获取模块,用于获取关于模板方案的参数配置限定信息,其中,所述参数配置限定信息用于生成参数配置界面,使得经由所述参数配置界面而配置的至少一个配置参数替换模板方案中的至少一个参数占位符,所述生成模块基于获取的模板方案、参数配置限定信息和输入源配置限定信息来生成机器学习方案模板的模板文件。

[0046] 可选地,所述参数配置限定信息包括以下项之中的至少一个:参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值。

[0047] 可选地,所述类型信息指示脚本参数和/或运行参数。

[0048] 可选地,所述参数配置限定信息还包括用于限定经由所述参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项。

[0049] 可选地,所述处理项包括对所述至少一个配置参数进行校验的校验项。

[0050] 可选地,所述第三获取模块:基于获取的模板方案来产生用于设置参数配置限定信息的控件;向第一用户展示产生的控件;以及接收第一用户通过所述控件所设置的参数配置限定信息。

[0051] 可选地,所述参数配置限定信息还包括:用于限定在参数配置界面上按照分类区域对参数占位符进行配置的分类信息,其中,所述第三获取模块还:向第一用户展示用于对参数占位符进行分类的控件;以及根据第一用户通过所述控件对参数占位符进行的分类来获取分类信息。

[0052] 可选地,该装置还包括:第一展示模块,用于向第一用户展示用于上传说明文档的控件;第一接收模块,用于接收第一用户通过所述控件所上传的说明文档;以及第一合并模块,用于将说明文档合并入模板文件。

[0053] 可选地,该装置还包括:第二展示模块,用于向第一用户展示用于设置资源配置信息的控件;第二接收模块,用于接收第一用户通过所述控件所设置的资源配置信息,所述资源配置信息用于表征执行所述至少部分机器学习过程的资源配置;以及第二合并模块,用于将资源配置合并入模板文件。

[0054] 可选地,该装置还包括:第三展示模块,用于向第三用户展示基于输入源配置限定信息生成的输入源配置界面;第四获取模块,用于获取第三用户基于与测试场景对应的测试数据表,经由输入源配置界面所配置的至少一个配置输入源;替换模块,用于用所述至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的机器学习方案模板;执行模块,用于基于修改后的机器学习方案模板执行所述至少部分机器学习过程,以得到所述至少部分机器学习过程的执行结果;评估模块,用于对所述执行结果进行评估,以得

到测试结果;发布或调试模块,用于基于所述测试结果确定是否发布所述机器学习方案模板,或者基于所述测试结果对所述机器学习方案模板进行调试。

[0055] 根据本发明的第四个方面,提出了一种基于机器学习方案模板执行机器学习过程的装置,包括:第一获取模块,用于获取机器学习方案模板的模板文件,其中,所述模板文件包括模板方案和输入源配置限定信息,所述模板方案用于描述针对至少一个输入源标记的至少部分机器学习过程,所述机器学习过程涉及模型训练和/或模型应用,所述输入源配置限定信息用于生成输入源配置界面;第一展示模块,用于向第二用户展示基于输入源配置限定信息而生成的输入源配置界面;第二获取模块,用于获取第二用户经由所述输入源配置界面而配置的至少一个配置输入源;第一替换模块,用于用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案;执行模块,用于基于修改后的模板方案来执行机器学习过程。

[0056] 可选地,所述输入源配置界面中包括用于设置配置输入源的控件,所述第二获取模块接收第二用户通过所述控件所设置的配置输入源。

[0057] 可选地,所述输入源标记用于标识模板方案中能够被替换的输入表和/或字段,所述输入源配置限定信息包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各第一字段的名称、各第一字段在输入源配置界面上是否展示为可选字段的指示信息,所述输入源配置界面中还包括以下项之中的至少一个:需要配置的至少一个输入表名称、各输入表下需要配置的各第一字段的名称、各第一字段是否为可选字段的指示信息,所述配置输入源包括以下项之中的至少一个:针对所述输入表配置的业务数据表、针对所述第一字段配置的业务数据表下的第二字段。

[0058] 可选地,所述输入源配置限定信息还包括各第一字段对应的字段格式,其中,至少一个第一字段对应的字段格式被设置为允许针对单个第一字段配置实际业务数据中的一个或多个第二字段,使得所配置的一个或多个第二字段均按照模板方案中处理所述单个第一字段的同样方式进行字段处理,所述输入源配置界面中用于设置针对所述第一字段的第二字段的控件是基于所述第一字段对应的字段格式生成的,以使得第二用户通过该控件能够按照第一字段对应的字段格式针对第一字段配置第二字段。

[0059] 可选地,所述输入源配置限定信息还包括用于限定经由所述输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项,该装置还包括:第一处理模块,用于按照所述处理项对所述配置输入源进行处理;第二展示模块,用于在所述输入源配置界面中展示处理结果。

[0060] 可选地,所述处理项包括关于各第一字段的校验项,其中,校验项包括各第一字段的允许格式和/或允许取值范围,按照所述处理项对所述配置输入源进行处理的步骤包括:按照所述第一字段的检验项,对所述第一字段配置的第二字段的格式和/或取值进行校验,其中,所述处理结果用于指示针对所述第一字段配置的第二字段的格式和/或取值是否符合所述校验项。

[0061] 可选地,校验项还包括是否进行校验的指示信息。

[0062] 可选地,所述模板方案包括至少一个参数占位符,所述模板文件还包括参数配置限定信息,所述参数配置限定信息用于生成参数配置界面,该装置还包括:第三展示模块,

用于向第二用户展示基于参数配置限定信息而生成的参数配置界面；第三获取模块，用于获取第二用户经由所述参数配置界面而配置的至少一个配置参数；第二替换模块，用于用获取的至少一个配置参数替换模板方案中的至少一个参数占位符，以得到修改后的模板方案。

[0063] 可选地，所述参数配置界面中包括用于设置配置参数的控件，所述第三获取模块接收第二用户通过所述控件所设置的配置参数。

[0064] 可选地，所述参数配置限定信息包括以下项之中的至少一个：参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值，所述参数配置界面中还包括以下项之中的至少一个：需要配置的参数占位符的类型信息、输入方式信息、展示信息、默认取值。

[0065] 可选地，所述参数配置限定信息还包括用于限定经由所述参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项，该装置还包括：第二处理模块，用于按照所述处理项对所述配置参数进行处理；第四展示模块，用于在所述输入源配置界面中展示处理结果。

[0066] 可选地，所述处理项包括对所述至少一个配置参数进行校验的校验项。

[0067] 可选地，所述类型信息指示脚本参数和/或运行参数。

[0068] 可选地，所述参数配置限定信息还包括：用于限定在参数配置界面上按照分类区域对参数占位符进行配置的分类信息，所述参数配置界面按照所述分类信息将需要配置的参数占位符进行分类显示，不同分类的参数占位符被显示在不同分类区域。

[0069] 可选地，所述模板文件还包括用于辅助第二用户了解和/或配置所述模板方案的说明文档，该装置还包括：提供模块，用于向所述第二用户提供所述说明文档。

[0070] 可选地，所述模板文件还包括资源配置信息，所述资源配置信息用于表征执行所述至少部分机器学习过程的资源配置，所述执行模块基于修改后的模板方案，使用所述资源配置信息所表征的资源配置执行所述机器学习过程，或者所述执行模块对修改后的模板方案在执行机器学习过程中所需的资源配置进行预测，使用预测得到的资源配置执行所述机器学习过程。

[0071] 根据本发明的第五个方面，提出了一种包括至少一个计算装置和至少一个存储指令的存储装置的系统，其中，指令在被至少一个计算装置运行时，促使至少一个计算装置执行如本发明第一个方面或第二个方面述及的方法。

[0072] 根据本发明的第六个方面，提出了一种存储指令的计算机可读存储介质，其中，当指令被至少一个计算装置运行时，促使至少一个计算装置执行如本发明第一个方面或第二个方面述及的方法。

[0073] 在根据本发明示例性实施例的机器学习方案模板的创建方法、使用方法及装置中，通过复用机器学习方案模板可以降低建模门槛并减少建模耗时，而机器学习方案模板的模板文件中的输入源配置限定信息可以用于解决实际业务数据与模板方案中固有数据之间的数据匹配问题，使得机器学习方案模板在应用于同一业务方向下不同数据结构的业务数据时，均能够获取较好的建模效果。

## 附图说明

[0074] 从下面结合附图对本发明实施例的详细描述中,本发明的这些和/或其他方面和优点将变得更加清楚并更容易理解,其中:

[0075] 图1示出了根据本发明示例性实施例的用于创建机器学习方案模板的方法的流程图;

[0076] 图2示出了机器学习方案模板的创建界面示意图;

[0077] 图3示出了根据本发明示例性实施例的基于机器学习方案模板执行机器学习过程的方法的流程图;

[0078] 图4示出了机器学习方案模板的配置界面示意图;

[0079] 图5示出了根据本发明示例性实施例的用于创建机器学习方案模板的装置的结构框图;

[0080] 图6示出了根据本发明示例性实施例的用于创建机器学习方案模板的装置的结构框图。

## 具体实施方式

[0081] 为了使本领域技术人员更好地理解本发明,下面结合附图和具体实施方式对本发明的示例性实施例作进一步详细说明。

[0082] 图1示出了根据本发明示例性实施例的用于创建机器学习方案模板的方法的流程图。图1所示的方法可完全通过计算机程序以软件方式实现,还可通过特定配置的计算装置来执行图1所示的方法。

[0083] 参见图1,在步骤S110,获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案。

[0084] 不同业务场景下的机器学习建模方案往往存在较大差别。例如时序场景中的机器学习建模方案注重时序窗口的搭建,对时序类型数据更为敏感,而营销场景中的机器学习建模方案更关注商品、用户标签等方面的数据。

[0085] 在本发明中,模板方案可以视为由用户(为了便于区分,可以称为第一用户)根据业务场景总结定义的一套该业务场景中通用的机器学习建模方案。其中,第一用户可以是指机器学习建模经验丰富的科学家。

[0086] 模板方案可以包括至少部分机器学习过程的各步骤配置。例如,模板方案可以描述至少部分机器学习过程中各步骤涉及的处理对象、处理方式、处理结果等一种或多种配置信息。

[0087] 机器学习过程涉及模型训练和/或模型应用。模型训练是指机器学习模型的训练过程,可以包括但不限于以下步骤之中的至少一个步骤:数据导入、数据拆分、特征抽取、模型训练、模型测试和模型评估,关于各步骤的详细描述可以参见现有机器学习知识,本发明不再赘述。模型应用是指机器学习模型的应用过程,例如可以是指使用训练好的机器学习模型对数据进行预测以得到预测结果的过程,作为示例,可包括打包应用、部署上线、提供服务等处理。

[0088] 模板方案可以是基于特定语言编写的用于描述至少部分机器学习过程的文件。例如,模板方案可以是DAG(有向无环图)文件,DAG文件可以描述有向无环图中各个节点(即下

文述及的处理节点)所表示的机器学习步骤的配置信息。

[0089] 模板方案中包括一个或多个输入源,输入源是指模板方案描述的机器学习过程所使用的输入源。其中,输入源可以包括但不限于输入表和/或字段,输入表也即模板方案描述的机器学习过程所使用的数据表,字段也即输入表中的字段。

[0090] 模板方案可以视为特定业务场景中通用的机器学习建模方案,然而同一业务场景下也可能采用数据结构不同的业务数据,即模板方案中的输入源和实际业务数据在数据结构上可能存在一定差异。为了使得模板方案中的输入源能够适应不同数据结构的业务数据,模板方案中的输入源可以是指能够被实际业务数据替换的输入源,关于替换的具体实现可以参见下文相关描述。

[0091] 为了便于区分,本发明使用输入源标记表征模板方案中能够被替换的输入源。输入源标记仅用于指代模板方案中能够被替换的输入源,关于输入源在模板方案中的限定形式,本发明不做限定。也即,模板方案中能够被替换的输入源可以使用特殊标记标识,也可以不用特殊标记标识。

[0092] 在步骤S120,获取关于模板方案的输入源配置限定信息。

[0093] 输入源配置限定信息用于生成输入源配置界面,使得经由输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记。其中,输入源配置界面是指面向使用机器学习方案模板的用户(为了便于区分,可以称为第二用户)展示的界面,用于辅助第二用户将实际业务数据与模板方案中的输入源标记对应起来。

[0094] 输入源配置限定信息可以由第一用户设置,第一用户可以针对模板方案中的一个或多个输入源设置输入源配置限定信息。本发明可以通过多种方式获取第一用户设置的输入源配置限定信息。例如,第一用户可以通过但不限于编辑文档的方式生成包括输入源配置限定信息的文件,本发明可以向第一用户提供文件上传接口,从第一用户通过该接口上传的文件中获取输入源配置限定信息。再例如,本发明也可以向第一用户提供可视化操作界面,根据第一用户在可视化操作界面上执行的操作,获取第一用户设置的输入源配置限定信息。

[0095] 作为示例,可以基于获取的模板方案来产生用于设置输入源配置限定信息的控件,向第一用户展示产生的控件,并接收第一用户通过控件所设置的输入源配置限定信息。以输入源指示输入表为例,在获取到模板方案中,可以对模板方案进行解析,以确定模板方案涉及的输入表,向第一用户提供用于添加输入表的控件,第一用户可以通过该控件添加输入表,并可以基于其他控件为该输入表设置相关的输入源配置限定信息。

[0096] 输入源配置限定信息可以包括任何能够用于辅助第二用户将实际业务数据与模板方案中的输入源对应起来的信息。以输入源标记用于标识模板方案中能够被替换的输入表和/或字段为例,输入源配置限定信息可以包括但不限于以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各字段的名称、各字段在输入源配置界面上是否展示为可选字段的指示信息。

[0097] 输入表名称是指向第二用户展示的表名称。第一用户可以将模板方案中输入表的名称作为向第二用户展示的表名称,也可以根据输入表的业务含义命名输入表名称,以便第二用户为输入表配置业务含义相同或相似的业务数据表。

[0098] 输入表对应的处理节点用于表征模板方案描述的机器学习过程中对输入表进行处理的节点,也即在哪个机器学习步骤处理输入表。作为示例,模板方案可以是上文提及的DAG文件,DAG文件中每个处理节点可以具有对应的节点ID,可以利用节点ID表征输入表对应的处理节点。

[0099] 字段的名称是指向第二用户展示的字段名。第一用户可以将输入表中字段的原始名称作为向第二用户展示的字段名,也可以根据字段的业务含义重新命名字段名,以便第二用户根据业务含义将业务数据表中的字段与输入表中的字段对应起来。需要说明的是,各输入表下需要配置的各字段,除了可以包括输入表中存在的原始字段,还可以包括输入表中不存在的字段,即扩展字段。举例来说,假设输入表A包括字段a、字段b和字段c,输入表A下需要配置的各字段除了可以包括字段a、字段b和字段c,还可以包括输入表A中不存在的字段d,字段d即为扩展字段。扩展字段可以由第一用户设置,例如第一用户可以从业务场景出发,根据业务场景中可能存在的业务数据结构,为输入表添加一个或多个输入表不存在的扩展字段,并设置扩展字段的字段名。其中,在添加扩展字段时,还可以设置扩展字段的处理方式。作为示例,为输入表添加的每个扩展字段还可以视为一个字段类别,不同扩展字段对应不同字段类别。通过设置扩展字段,可以为实际业务数据中超出模板方案涉及的字段的附加字段的匹配提供支持,即可以将附加字段与扩展字段对应起来,而扩展字段的处理方式又是预先设定好的,使得附加字段也可以参与机器学习,实现其数据价值,从而可以增强模板方案的数据适应性。

[0100] 指示信息用于指示字段是否为可选字段。可选字段是指第二用户可以根据实际情况决定是否为该字段配置实际业务数据中的字段。非可选字段,也即必选字段,是指第二用户需要为该字段配置实际业务数据中的至少一个字段。本发明可以根据业务场景中字段的特性(如通用性、重要性),将字段设置为可选字段或必选字段。例如第一用户可以将少量在业内通用的字段设置为必选字段,将其他字段设置为可选字段。其中,可选字段除了包括模板方案中存在的字段外,还可以包括模板方案中不存在的扩展字段。关于扩展字段可以参见上文相关描述,此处不再赘述。由此,通过将业内通用或重要性较高的字段划分为必选字段,可以保证建模效果不会太差,而通过划分可选字段,则可以优化模板方案在不同数据结构下的适应性。其中,至少一个字段对应的字段格式被设置为允许针对单个字段配置实际业务数据中的一个或多个字段,使得所配置的一个或多个字段均按照模板方案中处理单个字段的同样方式进行字段处理。由此,通过这种“一对多”的字段对应方式,使得即便实际业务数据中存在大量字段,也可以将这些字段与输入表中的字段对应起来,进而使得实际业务数据中的所有字段都可以参与到机器学习过程(如特征构造),实现其数据价值,并对最终结果产生影响。

[0101] 此外,输入源配置限定信息还可以包括各字段对应的字段格式。

[0102] 另外,输入源配置限定信息还可以包括用于限定经由输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项。其中,处理项可以包括但不限于关于各字段的校验项。校验项可以包括各字段的允许格式和/或允许取值范围,还可以包括是否进行校验的指示信息。

[0103] 在步骤S130,基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

[0104] 机器学习方案模板的模板文件包括模板方案和输入源配置限定信息。模板方案可以视为建模经验丰富的科学家根据业务场景总结定义的一套该业务场景中通用的机器学习建模方案,模板方案中包括科学家沉淀下来的一些机器学习建模know how,第二用户可以通过使用模板方案来复用这些建模knowhow,以降低建模门槛并减少建模耗时;输入源配置限定信息则可以用于辅助第二用户将实际业务数据与模板方案中输入源标记对应起来,以解决实际业务数据与模板方案中固有数据之间的数据匹配问题,使得模板方案在应用于同一业务方向下不同数据结构的业务数据时,均能够获取较好的建模效果。

[0105] 在生成模板文件后,还可以对模板文件进行测试,以判断模板文件能否达到预期,如果达到预期则可以发布机器学习方案模板,如果没有达到预期则可以对机器学习方案模板进行调试。

[0106] 作为示例,可以向第三用户展示基于输入源配置限定信息生成的输入源配置界面,第三用户可以是指测试人员;获取第三用户基于与测试场景对应的测试数据表,经由输入源配置界面所配置的至少一个配置输入源,这里配置输入源是指由第三用户根据测试数据针对模板方案中的输入源标记而配置的输入源,配置输入源可以包括针对输入表配置的测试数据表和/或针对输入表中的字段配置的测试数据表中的字段,具体配置过程可以参见下文结合图3的相关描述,此处不再赘述;用至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的机器学习方案模板;基于修改后的机器学习方案模板执行至少部分机器学习过程,以得到至少部分机器学习过程的执行结果;对执行结果进行评估,以得到测试结果,基于测试结果确定是否发布机器学习方案模板,或者基于测试结果对机器学习方案模板进行调试。例如如果测试结果符合预期则可以发布机器学习方案模板,如果测试结果不符合预期则可以对机器学习方案模板进行调试,如可以对模板文件中的输入源配置限定信息进行修改。

[0107] 模板方案还可以包括至少一个参数占位符,参数占位符可以是任何约定好的非编写代码常用的语言,如可以是“{\$占位符\$}”。参数占位符所表征的参数是指可以由第二用户确定的参数,可以包括但不限于脚本参数和/或运行参数。由此,第一用户在设置模板方案时,还可以根据经验对模板方案进行改造,对其中可以变化的参数以占位符的方式进行替换,例如可以将特征组合方式、超参数、运行资源等可以变化的参数设置为参数占位符。

[0108] 本发明还可以获取关于模板方案的参数配置限定信息。参数配置限定信息用于生成参数配置界面,使得经由参数配置界面而配置的至少一个配置参数替换模板方案中的至少一个参数占位符。其中,参数配置界面是指面向第二用户展示界面,用于辅助第二用户对参数占位符所表征的参数进行配置。

[0109] 参数配置限定信息可以由第一用户设置,第一用户可以针对模板方案中的各个参数占位符设置参数配置限定信息。本发明可以通过多种方式获取参数配置限定信息。例如,第一用户可以通过但不限于编辑文档的方式生成包括参数配置限定信息的文件,本发明可以向第一用户提供文件上传接口,从第一用户通过该接口上传的文件中获取参数配置限定信息。再例如,本发明也可以向第一用户提供可视化操作界面,根据第一用户在可视化操作界面上执行的操作,获取第一用户设置的参数配置限定信息。

[0110] 作为示例,可以基于获取的模板方案来产生用于设置参数配置限定信息的控件,向第一用户展示产生的控件,接收第一用户通过控件所设置的参数配置限定信息。例如,在

获取到模板方案后,可以对模板方案进行解析,识别模板方案中的参数占位符,并向第一用户展示用于设置与参数占位符相关的参数配置限定信息的控件,以便通过该控件获取第一用户设置的参数配置限定信息。

[0111] 参数配置限定信息可以包括任何能够用于辅助第二用户对参数占位符所表征的参数进行配置的信息。作为示例,参数配置限定信息可以包括但不限于以下项之中的至少一个:参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值。其中,类型信息可以指示脚本参数和/或运行参数;输入方式信息可以包括填写方式和/或选择方式,其中填写方式是指通过填写的方式进行输入,选择方式是指通过对所提供的多个选择项进行选择的方式进行输入;展示信息可以包括用于帮助第二用户了解需要输入的参数提示信息,如可以是指参数占位符所表征的参数的名称;默认取值可以是指输入方式为选择方式的情况下提供的选择项的展示值,实际取值可以是指所提供的选择项的实际值,展示值不同于实际值,其中展示值可以是指经过转译得到的便于用户理解的展示内容,由此在提供选择项时,还可以将选择性的实际值隐藏起来,而向用户暴露转译后的展示值。如此既方便用户理解,同时还可以将第一用户设置的实际值作为技术诀窍(know how)保护起来。

[0112] 参数配置限定信息还可以包括用于限定经由参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理项。其中,处理项可以包括但不限于对至少一个配置参数进行校验的校验项。其中,校验项可以包括配置参数的允许格式和/或允许取值范围,还可以包括是否进行校验的指示信息。

[0113] 在模板方案中包括参数配置限定信息时,本发明可以基于获取的模板方案、参数配置限定信息和输入源配置限定信息来生成机器学习方案模板的模板文件。即模板文件中不仅可以包括模板方案和输入源配置限定信息,还可以包括参数配置限定信息。参数配置限定信息可以辅助第二用户对模板方案中的参数占位符所表征的参数进行配置,以保证整个建模效果与实际业务场景相适应,提升模板方案在实际业务场景中的应用效果。

[0114] 参数配置限定信息还可以包括用于限定在参数配置界面上按照分类区域对参数占位符进行分类的信息。在获取关于模板方案的参数配置限定信息的过程中,本发明还可以向第一用户展示用于对参数占位符进行分类的控件,根据第一用户通过控件对参数占位符进行的分类来获取分类信息。由此,所生成的参数配置界面可以按照分类信息将需要配置参数占位符进行分类显示,其中不同分类的参数占位符被显示在不同分类区域,使得第二用户对参数占位符进行配置时更加有逻辑感。

[0115] 本发明还可以向第一用户展示用于上传说明文档的控件,接收第一用户通过控件所上传的说明文档,并将说明文档合并入模板文件。说明文档可以是用于告知用户如何对模板方案进行配置的文档,该文档可以用便于第二用户理解的业务语言描述的,使得第二用户在大部分情况下不需要理解机器学习的概念即可使用机器学习方案模板。例如,说明文档可以包括样本标签的设置说明,使得用户不需要理解机器学习的正样本、负样本概念,只需根据业务如实反应情况,在反欺诈业务中,用户只需明确哪些是问题交易,哪些是正常交易,通过正常业务理解的方式即可进行整个建模模板的操作,降低了模板使用的门槛。

[0116] 本发明还可以向第一用户展示用于设置资源配置信息的控件,接收第一用户通过

控件所设置的资源配置信息,资源配置信息用于表征执行至少部分机器学习过程的资源配置,将资源配置合并入模板文件。此处述及的资源配置信息可以视为由第一用户提供的运行资源设置。

[0117] 图1所示的方法可以由用于实现机器学习相关业务的机器学习平台执行。图2示出了由机器学习平台向用户展示的机器学习方案模板的创建界面示意图。此处述及的用户是指用于创建机器学习方案模板的用户,也即上文述及的第一用户,例如可以是机器学习建模经验丰富的科学家。

[0118] 如图2所示,机器学习方案模板的创建可以分为四部分,分别是基础信息配置、输入配置、占位符配置以及分类配置。

[0119] 1、基础信息配置

[0120] 基础信息可以包括但不限于建模模板名称、建模模板可见状态、建模模板DAG、建模模板配置说明文档。

[0121] 建模模板名称是指展示给用户(模板使用者,也即上文述及的第二用户)查看的机器学习方案模板的名称。科学家可以根据机器学习方案模板适用的业务场景、机器学习方案模板的功能进行命名,例如科学家可以将针对营销场景创建的机器学习方案模板命名为通用营销建模模板。

[0122] 建模模板可见状态是指建模模板是否对其他用户可见。如选择不可见则用户无法在前台查看,反之用户可以在前台入口查看。

[0123] 建模模板DAG是指机器学习方案模板的DAG文件。科学家可以通过点击上传控件上传DAG文件,上传完成后后台可以自动扫描DAG文件中的占位符信息,基于扫描结果在界面中展示用于科学家对占位符进行配置的控件。

[0124] 建模模板配置说明文档用于告知用户如何对该建模模板进行配置的文档,可以包含每个配置字段的详细说明,帮助用户理解。

[0125] 2、输入配置

[0126] 输入配置是指由科学家设置输入源配置限定信息。

[0127] 初始情况下,界面中可以仅显示一个“添加输入表”控件,第一用户点击“添加输入表”控件后,界面中可以显示一个输入表模块信息。

[0128] 输入表模块信息包括用于填写输入表名称的控件、用于填写节点ID(即图中示出的node ID)的控件、用于添加字段的控件以及字段展示表。字段展示表中包括用于填写字段名的控件、用于设置是否需要字段校验的控件以及用于选择字段类型的控件。

[0129] 科学家可以根据输入表的业务含义设置输入表名称,以便用户可以根据输入表名称选择合适的实际业务数据中的数据表进行匹配。相应地,科学家也可以根据字段含义重命名字段名以保障用户看到的字段名易于理解。科学家还可以设置是否需要字段校验,如果选择需要字段校验,还需要设置字段类型,便于校验用户提供的字段类型与科学家要求是否一致。

[0130] 3、占位符配置

[0131] 可以读取科学家上传的DAG中的占位符并展示对应的占位符模块,每个占位符模块包含一个选项框、一个占位符名称和与选项框对应的选项内容。占位符变量有脚本变量和运行参数两种。

[0132] 占位符选择脚本变量时,选项框包含选择项和输入项两种选择。选择项代表用户会看到一个选择类的填写需求,选择类只支持单选,并设置对应的选项,选项最少两个,可点击添加选项来增加选项;填写项代表用户会看到一个含有填写项名称的文本框,用户填写内容即可。

[0133] 选项名用于帮助用户了解该选项需要填写的信息,选项名的设置应尽量与建模模板配置说明文档一致。如选择校验字段类型,需根据科学家设置字段类型进行填写内容校验,如不选择校验字段类型则无需校验字段。字段类型是指科学家设置的字段类型,便于校验用户提供的字段类型与科学家要求是否一致,可以包含但不限于enums、String。如选择校验字段阈值,需根据字段阈值校验,如不选择,则不校验用户填写的数据信息。字段阈值可填可不填,例如可填写正则表达式、闭合区间(代表大于等于,小于等于的区间)表征字段阈值。填写后会校验对应表单字段列下的数据是否符合区间要求,如填写为正则表达式,还可以提供正则表达式提示文案,主要内容为说明需要填写什么字段信息。

[0134] 占位符选择运行参数时,只允许用户填写,科学家可以通过节点别名(节点名称,例如可以是节点ID)指定算子并明确参数的名称(代码内英文名),设置显示给用户填写的参数名、默认的参数项、参数字段类型和对应的阈值正则表达式及正则表达式的提示文案。运行参数字段类型可以支持但不限于String、Int、Double、Boolean、Long。

#### [0135] 4、分类配置

[0136] 在每一个占位符变量下有一个添加分类线按钮。分类线以上至上一设置的分类线在展示给用户时作为一个模块展示,如分类线以上没有其他分类线模块设置,则该分类线到第一个占位符变量组成一个模块展示。

[0137] 科学家需要输入分类名称以用于分类展示,分类名称用于表征作为一个模块展示的占位符变量的类别信息。以两个分类线之间的占位符变量分别为营销时序窗口配置和理财购买时序窗口配置为例,科学家可以将分类名称设置为时序参数设置。

[0138] 科学家在完成设置后点击保存方案控件,平台会保存当前建模模板。科学家还可以在建模模板管理页面查看所创建的建模模板。其中,创建时间为点击保存方案的时间。科学家点击取消可以发出“取消后将不保留当前所有设置,请谨慎取消”的提醒消息。

[0139] 至此结合图1、图2就本发明的创建机器学习方案模板的方法的流程做了详细说明。本发明还提出了一种基于机器学习方案模板执行机器学习过程的方法的流程图。其中机器学习方案模板可以是基于本发明的创建机器学习方案模板的方法生成的。因此,本发明的基于机器学习方案模板执行机器学习过程的方法还可以包括图1所示的各步骤。

[0140] 图3示出了根据本发明示例性实施例的基于机器学习方案模板执行机器学习过程的方法的流程图。图3所示的方法可完全通过计算机程序以软件方式实现,还可通过特定配置的计算装置来执行图3所示的方法。

[0141] 参见图3,在步骤S310,获取机器学习方案模板的模板文件。

[0142] 模板文件包括模板方案和输入源配置限定信息。模板方案用于描述针对至少一个输入源标记的至少部分机器学习过程,机器学习过程涉及模型训练和/或模型应用,输入源配置限定信息用于生成输入源配置界面。关于模板方案、输入源标记、输入源配置限定信息可以参见上文相关描述,此处不再赘述。

[0143] 在步骤S320,向第二用户展示基于输入源配置限定信息而生成的输入源配置界

面。输入源配置界面用于辅助第二用户将实际业务数据与模板方案中输入源标记对应起来。其中,第二用户是指使用机器学习方案模板的用户。第二用户可以是机器学习建模经验丰富的科学家,也可以是机器学习建模经验欠缺的业务人员。

[0144] 在步骤S330,获取第二用户经由输入源配置界面而配置的至少一个配置输入源。输入源配置界面中可以包括用于设置配置输入源的控件,可以接收第二用户通过控件所设置的配置输入源。

[0145] 配置输入源来自于实际业务数据。配置输入源是指由第二用户根据实际业务数据针对模板方案中的输入源标记而配置的输入源。以输入源标记用于标识模板方案中能够被替换的输入表和/或字段为例,模板方案中输入表下的字段可以称为第一字段,业务数据表下的字段可以称为第二字段,配置输入源可以包括针对输入表配置的业务数据表和/或针对第一字段配置的业务数据表下的第二字段。其中,业务数据表是指实际业务场景中生成的数据表,业务数据表表征的是实际业务数据。

[0146] 以输入源标记用于标识模板方案中能够被替换的输入表和/或字段为例,输入源配置限定信息可以包括以下项之中的至少一个:输入源配置界面上展示的需要配置的至少一个输入表名称、各输入表对应的处理节点、各输入表下需要配置的各第一字段的名称、各第一字段在输入源配置界面上是否展示为可选字段的指示信息。输入源配置界面中还可以包括以下项之中的至少一个:需要配置的至少一个输入表名称、各输入表下需要配置的各第一字段的名称、各第一字段是否为可选字段的指示信息。

[0147] 作为示例,第二用户首先可以根据输入源配置界面上展示的输入表名称,配置对应的业务数据表,然后根据各输入表下需要配置的各第一字段的名称及其是否为可选字段的指示信息,为第一字段配置对应的第二字段,以将该业务数据表中的第二字段与输入表中的第一字段对应起来。

[0148] 对于属于可选字段的第一字段,第二用户可以判断业务数据表中是否存在与该第一字段匹配的第二字段,如果存在匹配的第二字段,则为该第一字段设置对应的第二字段,如果不存在匹配的第二字段,则可以不为该第一字段设置第二字段。对于不属于可选字段(也即属于必选字段)的第一字段,第二用户需要为该字段配置业务数据表中至少一个第二字段。

[0149] 如上文所述,第一字段是否为可选字段,可以是根据业务场景中字段的通用性或重要性确定的,如可以将业务通用或重要性较高的字段设置为必选字段,将业务不通用或重要性不高的字段设置为可选字段。由此,第二用户在将业务数据表中的第二字段和输入表中的第一字段进行匹配时,属于必选字段的第一字段至少存在一个对应的第二字段,如此可以保证模板方案应用到实际业务场景中时建模效果不会太差,而属于可选字段的第一字段是否存在对应的第二字段是由第二用户根据实际情况设定的,即第二用户无需为输入表下各个第一字段都配置对应的第二字段,使得在保证建模效果的同时还可以优化数据表的适应性。

[0150] 如上文所述,各输入表下需要配置的第一字段除了可以包括输入表中存在的原始字段,还可以包括输入表中不存在的字段,即扩展字段。扩展字段可以是根据业务场景中可能存在的数据结构为输入表添加的,并且所添加的扩展字段还可以具有对应的处理方式。因此第二用户在将业务数据表中的第二字段和输入表中的第一字段进行匹配时,还可以将

第二字段与输入表中不存在的字段(即扩展字段)对应起来。由此,即使业务数据表中的第二字段的数量较多,超出了模板方案中输入表下的字段,第二用户也可以通过将超出的字段与扩展字段对应起来,使得这些超出的字段也可以参与机器学习,实现数据价值,并对最终结果产生影响。例如上文所述,为输入表添加的每个扩展字段还可以视为一个字段类别,不同扩展字段对应不同字段类别,对于业务数据表中超出输入表涉及的字段的附加字段,第二用户可以将附加字段划分到对应类别的扩展字段下,使得附加字段也可以参与机器学习,实现其数据价值。

[0151] 例如,至少一个第一字段对应的字段格式被设置为允许针对单个第一字段配置实际业务数据中的一个或多个第二字段,使得所配置的一个或多个第二字段均按照模板方案中处理单个第一字段的同样方式进行字段处理,输入源配置界面中用于设置针对第一字段的第二字段的控件是基于第一字段对应的字段格式生成的,以使得第二用户通过该控件能够按照第一字段对应的字段格式针对第一字段配置第二字段。

[0152] 也就是说,对于输入源配置界面中输入表下需要配置的第一字段,该第一字段的字段格式可以是只允许针对该字段配置一个第二字段,或者也可以是允许针对该字段配置多个第二字段。也即,输入源配置界面中输入表下需要配置的第一字段中可以既可以包括支持“一对一”配置的第一字段,也可以包括支持“一对多”配置的第一字段。对于支持“一对一”配置的第一字段,允许第二用户为该字段配置最多一个第二字段,对于支持“一对多”配置的第一字段,允许第二用户为该字段配置多个第二字段。

[0153] 由此,即便业务数据表中存在大量字段,通过这种“一对多”的字段对应方式,也可以将这些字段与输入表中的第一字段对应起来,进而使得业务数据表中的所有字段都可以参与到建模过程,实现其数据价值,并对最终建模结果产生影响。

[0154] 作为本发明的一个示例,输入源配置界面中展示的输入表可以包括至少一个可选表单,该可选表单中可以包括多个列名。可选表单中的列名可以是模板方案中已有的基本字段(例如卡号、开户时间等),这些基本字段在模板方案中已经具有相应的处理方式。针对实际业务数据中超出模板方案涉及的字段的附加字段,第二用户可将这样的附加字段填写到可选表单对应的列名之下。比如,附加字段为激活时间,则可将该激活时间的字段名称填写到“开户时间”之下,使得可以按照开户时间的处理方式来处理激活时间。

[0155] 可选表单中的列名也可以是字段类别(例如可以是上文述及的扩展字段),其中每个字段类别具有对应的处理方式。针对模板方案中不涉及的附加字段,第二用户可将这样的附加字段填写到可选表单对应的列名之下。比如,附加字段为激活时间,则可将该激活时间的字段名称填写到“时间”类别之下,使得可以按照“时间”类别的处理方式来处理激活时间。

[0156] 此外,输入源配置限定信息还可以包括各第一字段对应的字段格式。

[0157] 另外,输入源配置限定信息还可以包括用于限定经由输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项。本发明还可以按照处理项对配置输入源进行处理,并在输入源配置界面中展示处理结果。

[0158] 作为示例,处理项可以包括关于各第一字段的校验项,校验项可以包括各第一字段的允许格式和/或允许取值范围。此时可以按照第一字段的检验项,对第一字段配置的第二字段的格式和/或取值进行校验,处理结果用于指示针对第一字段配置的第二字段的格

式和/或取值是否符合校验项。可选地,校验项还可以包括是否进行校验的指示信息。

[0159] 在步骤S340,用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案。

[0160] 以配置输入源包括针对输入表配置的业务数据表和针对第一字段配置的业务数据表下的第二字段为例,可以将模板方案中的输入表替换为所配置的业务数据表。输入表及输入表下的第一字段在模板方案中的处理方式是已知的,在将配置输入源替换掉模板方案中的输入源标记时,可以根据输入表的处理方式设定替换后的业务数据表的处理方式,根据第一字段的处理方式设定为其配置的第二字段的处理方式。由此,修改后的模板方案描述的是针对配置输入源的机器学习过程。

[0161] 由此,第二用户在使用机器学习方案模板时,不需要根据模板方案的字段要求而产生对应的表格,只需要将相同业务含义的字段做匹配,就可以完成实际业务数据的引入,避免了用户对数据结构的反复修改工作。

[0162] 在步骤S350,基于修改后的模板方案来执行机器学习过程。

[0163] 修改后的模板方案描述的是针对配置输入源的机器学习过程,而配置输入源表征的则是实际业务数据。因此基于修改后的模板方案执行机器学习过程,可以得到符合实际业务场景的机器学习结果。

[0164] 综上,模板方案可以视为建模经验丰富的科学家根据业务场景总结定义的一套该业务场景中通用的机器学习建模方案,模板方案中包括科学家沉淀下来的一些机器学习建模know how,第二用户可以通过使用机器学习建模方案来复用这些建模know how,以降低建模门槛并减少建模耗时;并且在使用模板的过程中,第二用户不需要理解机器学习方案模板的建模原理,也不需要了解建模过程,而仅通过可视化地方式将实际业务数据与模板方案中的数据对应起来,就可以得到符合业务场景的机器学习结果。

[0165] 如上文所述,模板方案还可以包括至少一个参数占位符,模板文件还包括用于生成参数配置界面的参数配置限定信息。关于参数占位符、参数配置界面、参数配置限定信息可以参见上文相关描述,此处不再赘述。

[0166] 本发明还可以向第二用户展示基于参数配置限定信息而生成的参数配置界面,获取第二用户经由参数配置界面而配置的至少一个配置参数,用获取的至少一个配置参数替换模板方案中的至少一个参数占位符,以得到修改后的模板方案。由此,可以向第二用户开放一些可配置的参数,这些可配置的参数可以是需要根据业务要求调整的部分,通过这种方式,可以保证机器学习过程的执行效果与实际业务场景相适应,提升机器学习方案模板的使用效果。

[0167] 参数配置界面中可以包括用于设置配置参数的控件,可以接收第二用户通过控件所设置的配置参数。

[0168] 参数配置限定信息还可以包括但不限于以下项之中的至少一个:参数配置界面上展示的需要针对参数占位符进行配置的类型信息、输入方式信息、展示信息、默认取值、实际取值。参数配置界面中还可以包括但不限于以下项之中的至少一个:需要配置的参数占位符的类型信息、输入方式信息、展示信息、默认取值。关于类型信息、输入方式信息、展示信息、默认取值、实际取值可以参见上文相关描述,此处不再赘述。

[0169] 参数配置限定信息还可以包括用于限定经由参数配置界面而配置的至少一个配

置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项,本发明还可以按照处理项对配置参数进行处理,并在输入源配置界面中展示处理结果。处理项可以包括对至少一个配置参数进行校验的校验项。其中,校验项可以包括但不限于配置参数的允许格式和/或允许取值范围,还可以包括是否进行校验的指示信息。

[0170] 参数配置限定信息还可以包括用于限定在参数配置界面上按照分类区域对参数占位符进行分类信息的分类信息,参数配置界面可以按照分类信息将需要配置的参数占位符进行分类显示,不同分类的参数占位符被显示在不同分类区域。分类显示也即属于同一类别的参数作为一个分组进行显示,使得用户填写的时候更加有逻辑感。

[0171] 模板文件还可以包括用于辅助第二用户了解和/或配置模板方案的说明文档,本发明还可以向第二用户提供说明文档。说明文档可以是用便于第二用户理解的业务语言描述的,使得第二用户在大部分情况下不需要理解机器学习的概念即可使用机器学习方案模板。

[0172] 作为本发明的一个示例,模板文件还可以包括资源配置信息,资源配置信息用于表征执行至少部分机器学习过程的资源配置,其中资源配置信息可以是由第一用户设置的。在执行步骤S350时,可以基于修改后的模板方案,使用资源配置信息所表征的资源配置执行机器学习过程。

[0173] 作为本发明的另一个示例,在执行步骤S350时,也可以对修改后的模板方案在执行机器学习过程中所需的资源配置进行预测,使用预测得到的资源配置执行机器学习过程。其中,可以采用但不限规则公式、预测模型的方式对所需的资源配置进行预测。例如可以根据规则公式,结合样本数据量、特征抽取方式、模型训练算法等推算所需的资源配置,再例如也可以提供建模任务的各种数据量在不同资源配置(例如,从小到大递增的资源配置)下的试跑,将试跑成功和失败时的数据量及资源配置作为样本,预先训练用于推测所需资源配置的机器学习模型,使用机器学习模型对所需的资源进行预测。

[0174] 图3所示的方法可以由用于实现机器学习相关业务的机器学习平台执行。图4示出了由机器学习平台向用户展示的机器学习方案模板的配置界面示意图。此处述及的用户是指机器学习方案模板的使用者,也即上文述及的第二用户。

[0175] 如图4所示,左上角显示的是建模模板的名称,建模模板配置说明为第一用户创建该建模模板时上传的说明文档。通过建模模板配置说明可了解当前建模模板需要配置哪些数据表、哪些字段、相应业务含义。

[0176] 用户表、商品表、行为表为模板方案中需要配置的输入表的名称,用户需填写对应的业务数据表,用户可直接填写业务数据表的表名称,在用户填写过程中可以搜索与用户输入的内容相匹配的业务数据表,并由下拉框弹出匹配项,匹配项可以按照时间倒序排列,每页显示5条业务数据表,用户可通过下滑操作查看。

[0177] 以用户表为例,user\_id、age、sex等为用户表下需要配置的字段。用户在为用户表选好对应的业务数据表后,可以针对用户表下需要配置的字段,选择业务数据表中与其对应的字段。字段填写完成后,可以校验所填字段是否符合要求,可以对字段类型、字段阈值进行校验,如不符合字段类型要求,可以提醒“#字段名#要求输入string字段,当前配置字段不符合要求请重新配置”,对应错误字段显示红星;如不符合字段阈值要求,可以提醒“#字段名#只接受#阈值下限#-#阈值上限#的数据,请检查对应表数据”。

[0178] 时序参数设置、样本参数设置是指需要配置的参数占位符。可以根据参数配置限定信息展示对应的名称和输入内容。如有参数配置限定信息包括设置的一个或多个选择项,则展示选择项,供用户选择。作为示例,可以展示可供用户调整的默认值。

[0179] 用户填写完成后,可以根据用户填写的内容,对模板方案中的可替换部分(如输入源标记、参数占位符)进行替换,使用替换后的模板方案执行机器学习过程。

[0180] 本发明的用于创建机器学习方案模板的方法,还可以实现为一种用于创建机器学习方案模板的装置。图5示出了根据本发明示例性实施例的用于创建机器学习方案模板的装置的结构框图。其中,用于创建机器学习方案模板的装置的功能单元可以由实现本发明原理的硬件、软件或硬件和软件的结合来实现。本领域技术人员可以理解的是,图5所描述的功能单元可以组合起来或者划分成子单元,从而实现上述发明的原理。因此,本文的描述可以支持对本文描述的功能单元的任何可能的组合、或者划分、或者更进一步的限定。

[0181] 下面就用于创建机器学习方案模板的装置可以具有的功能单元以及各功能单元可以执行的操作做简要说明,对于其中涉及的细节部分可以参见上文相关描述,这里不再赘述。

[0182] 参见图5,用于创建机器学习方案模板的装置500包括第一获取模块510、第二获取模块520以及生成模块530。

[0183] 第一获取模块510用于获取用于描述针对至少一个输入源标记的至少部分机器学习过程的模板方案,其中,机器学习过程涉及模型训练和/或模型应用。

[0184] 第二获取模块520用于获取关于模板方案的输入源配置限定信息,其中,输入源配置限定信息用于生成输入源配置界面,使得经由输入源配置界面而配置的至少一个配置输入源替换模板方案中的至少一个输入源标记。

[0185] 作为示例,第二获取模块520可以基于获取的模板方案来产生用于设置输入源配置限定信息的控件,向第一用户展示产生的控件,接收第一用户通过所述控件所设置的输入源配置限定信息。

[0186] 关于模板方案、输入源标记、输入源配置限定信息、输入源配置界面可以参见上文相关描述,此处不再赘述。

[0187] 生成模块530用于基于获取的模板方案和输入源配置限定信息生成机器学习方案模板的模板文件。

[0188] 模板方案中可以包括至少一个参数占位符,用于创建机器学习方案模板的装置500还可以包括第三获取模块。第三获取模块用于获取关于模板方案的参数配置限定信息,其中,参数配置限定信息用于生成参数配置界面,使得经由参数配置界面而配置的至少一个配置参数替换模板方案中的至少一个参数占位符。生成模块530可以基于获取的模板方案、参数配置限定信息和输入源配置限定信息来生成机器学习方案模板的模板文件。关于参数配置限定信息、参数配置界面可以参见上文相关描述,此处不再赘述。

[0189] 作为示例,第三获取模块基于获取的模板方案来产生用于设置参数配置限定信息的控件,向第一用户展示产生的控件,接收第一用户通过控件所设置的参数配置限定信息。参数配置限定信息还可以包括用于限定在参数配置界面上按照分类区域对参数占位符进行配置的分类信息。所述第三获取模块还向第一用户展示用于对参数占位符进行分类的控件,根据第一用户通过所述控件对参数占位符进行的分类来获取分类信息。

[0190] 用于创建机器学习方案模板的装置500还可以包括第一展示模块、第一接收模块以及第一合并模块。第一展示模块用于向第一用户展示用于上传说明文档的控件；第一接收模块用于接收第一用户通过控件所上传的说明文档；第一合并模块用于将说明文档合并入模板文件。

[0191] 用于创建机器学习方案模板的装置500还可以包括第二展示模块、第二接收模块以及第二合并模块。第二展示模块用于向第一用户展示用于设置资源配置信息的控件；第二接收模块用于接收第一用户通过控件所设置的资源配置信息，资源配置信息用于表征执行至少部分机器学习过程的资源配置；第二合并模块用于将资源配置合并入模板文件。

[0192] 用于创建机器学习方案模板的装置500还可以包括第三展示模块、第四获取模块、替换模块、执行模块、评估模块以及发布或调试模块。第三展示模块用于向第三用户展示基于输入源配置限定信息生成的输入源配置界面；第四获取模块用于获取第三用户基于与测试场景对应的测试数据表，经由输入源配置界面所配置的至少一个配置输入源；替换模块用于用所述至少一个配置输入源替换模板方案中的至少一个输入源标记，以得到修改后的机器学习方案模板；执行模块用于基于修改后的机器学习方案模板执行所述至少部分机器学习过程，以得到所述至少部分机器学习过程的执行结果；评估模块用于对所述执行结果进行评估，以得到测试结果；发布或调试模块用于基于所述测试结果确定是否发布所述机器学习方案模板，或者基于所述测试结果对所述机器学习方案模板进行调试。

[0193] 应该理解，根据本发明示例性实施例的用于创建机器学习方案模板的装置500的具体实现方式可参照结合图1、图2针对用于创建机器学习方案模板的方法的相关描述来实现，在此不再赘述。

[0194] 本发明的基于机器学习方案模板执行机器学习过程的方法，还可以实现为一种基于机器学习方案模板执行机器学习过程的装置。图6示出了根据本发明示例性实施例的用于创建机器学习方案模板的装置的结构框图。其中，基于机器学习方案模板执行机器学习过程的装置的功能单元可以由实现本发明原理的硬件、软件或硬件和软件的结合来实现。本领域技术人员可以理解的是，图6所描述的功能单元可以组合起来或者划分成子单元，从而实现上述发明的原理。因此，本文的描述可以支持对本文描述的功能单元的任何可能的组合、或者划分、或者更进一步的限定。

[0195] 下面就基于机器学习方案模板执行机器学习过程的装置可以具有的功能单元以及各功能单元可以执行的操作做简要说明，对于其中涉及的细节部分可以参见上文相关描述，这里不再赘述。

[0196] 参见图6，基于机器学习方案模板执行机器学习过程的装置600包括第一获取模块610、第一展示模块620、第二获取模块630、第一替换模块640以及执行模块650。

[0197] 第一获取模块610用于获取机器学习方案模板的模板文件，其中，模板文件包括模板方案和输入源配置限定信息，模板方案用于描述针对至少一个输入源标记的至少部分机器学习过程，机器学习过程涉及模型训练和/或模型应用，输入源配置限定信息用于生成输入源配置界面。关于模板方案、输入源标记、输入源配置限定信息可以参见上文相关描述，此处不再赘述。

[0198] 第一展示模块620用于向第二用户展示基于输入源配置限定信息而生成的输入源配置界面。关于输入源配置界面可以参见上文相关描述，此处不再赘述。

[0199] 第二获取模块630用于获取第二用户经由输入源配置界面而配置的至少一个配置输入源。输入源配置界面中可以包括用于设置配置输入源的控件,第二获取模块630可以接收第二用户通过控件所设置的配置输入源。关于配置输入源可以参见上文相关描述,此处不再赘述。

[0200] 第一替换模块640用于用获取的至少一个配置输入源替换模板方案中的至少一个输入源标记,以得到修改后的模板方案。

[0201] 执行模块650用于基于修改后的模板方案来执行机器学习过程。模板文件还可以包括资源配置信息,资源配置信息用于表征执行至少部分机器学习过程的资源配置,执行模块650可以基于修改后的模板方案,使用资源配置信息所表征的资源配置执行机器学习过程。或者执行模块650可以对修改后的模板方案在执行机器学习过程中所需的资源配置进行预测,使用预测得到的资源配置执行机器学习过程。执行模块650的具体实现细节可以参见上文相关描述

[0202] 输入源配置限定信息还包括用于限定经由输入源配置界面而配置的至少一个配置输入源在替换模板方案中的至少一个输入源标记之前所经过的处理的处理项,基于机器学习方案模板执行机器学习过程的装置600还可以包括第一处理模块和第二展示模块。第一处理模块用于按照处理项对配置输入源进行处理;第二展示模块用于在输入源配置界面中展示处理结果。关于处理项可以参见上文相关描述,此处不再赘述。

[0203] 模板方案中还可以包括至少一个参数占位符,模板文件还可以包括参数配置限定信息,参数配置限定信息用于生成参数配置界面,基于机器学习方案模板执行机器学习过程的装置600还可以包括第三展示模块、第三获取模块以及第二替换模块。第三展示模块用于向第二用户展示基于参数配置限定信息而生成的参数配置界面;第三获取模块用于获取第二用户经由参数配置界面而配置的至少一个配置参数;第二替换模块用于用获取的至少一个配置参数替换模板方案中的至少一个参数占位符,以得到修改后的模板方案。参数配置界面中可以包括用于设置配置参数的控件,第三获取模块可以接收第二用户通过控件所设置的配置参数。关于参数配置限定信息、参数配置界面可以参见上文相关描述,此处不再赘述。

[0204] 参数配置限定信息还可以包括用于限定经由参数配置界面而配置的至少一个配置参数在替换模板方案中的至少一个参数占位符之前所经过的处理的处理项,基于机器学习方案模板执行机器学习过程的装置600还可以包括第二处理模块和第四展示模块。第二处理模块用于按照处理项对配置参数进行处理;第四展示模块用于在输入源配置界面中展示处理结果。关于处理项可以参见上文相关描述,此处不再赘述。

[0205] 模板文件还包括用于辅助第二用户了解和/或配置模板方案的说明文档,基于机器学习方案模板执行机器学习过程的装置600还可以包括提供模块,提供模块用于向第二用户提供说明文档。

[0206] 应该理解,根据本发明示例性实施例的基于机器学习方案模板执行机器学习过程的装置600的具体实现方式可参照上文结合图3、图4针对基于机器学习方案模板执行机器学习过程的方法的相关描述来实现,在此不再赘述。

[0207] 以上参照图1到图6描述了根据本发明示例性实施例的机器学习方案模板的创建方法、使用方法及装置。应理解,上述方法可通过记录在计算可读介质上的程序来实现,例

如,根据本发明的示例性实施例,可提供一种存储指令的计算机可读存储介质,其中,在所述计算机可读介质上记录有用于执行本发明的用于创建机器学习方案模板的方法(例如图1所示)或基于机器学习方案模板执行机器学习过程的方法(例如图3所示)的计算机程序。

[0208] 上述计算机可读介质中的计算机程序可在诸如客户端、主机、代理装置、服务器等计算机设备中部署的环境中运行,应注意,所述计算机程序除了可用于执行除了图1或图3示出的步骤之外,还可用于执行除了上述步骤以外的附加步骤或者在执行上述步骤时执行更为具体的处理,这些附加步骤和进一步处理的内容已经参照图1、图3进行了描述,这里为了避免重复将不再进行赘述。

[0209] 应注意,根据本发明示例性实施例的用于创建机器学习方案模板的装置和基于机器学习方案模板执行机器学习过程的装置,可完全依赖计算机程序的运行来实现相应的功能,即,各个装置与计算机程序的功能架构中与各步骤相应,使得整个装置通过专门的软件包(例如,lib库)而被调用,以实现相应的功能。

[0210] 另一方面,图5、图6所示的各个装置也可以通过硬件、软件、固件、中间件、微代码或其任意组合来实现。当以软件、固件、中间件或微代码实现时,用于执行相应操作的程序代码或者代码段可以存储在诸如存储介质的计算机可读介质中,使得处理器可通过读取并运行相应的程序代码或者代码段来执行相应的操作。

[0211] 例如,本发明的示例性实施例还可以实现为计算装置,该计算装置包括存储部件和处理器,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行用于创建机器学习方案模板的方法或基于机器学习方案模板执行机器学习过程的方法。

[0212] 具体说来,所述计算装置可以部署在服务器或客户端中,也可以部署在分布式网络环境中的节点装置上。此外,所述计算装置可以是PC计算机、平板装置、个人数字助理、智能手机、web应用或其他能够执行上述指令集合的装置。

[0213] 这里,所述计算装置并非必须是单个的计算装置,还可以是任何能够单独或联合执行上述指令(或指令集)的装置或电路的集合体。计算装置还可以是集成控制系统或系统管理器的一部分,或者可被配置为与本地或远程(例如,经由无线传输)以接口互联的便携式电子装置。

[0214] 在所述计算装置中,处理器可包括中央处理器(CPU)、图形处理器(GPU)、可编程逻辑装置、专用处理器系统、微控制器或微处理器。作为示例而非限制,处理器还可包括模拟处理器、数字处理器、微处理器、多核处理器、处理器阵列、网络处理器等。

[0215] 根据本发明示例性实施例的用于创建机器学习方案模板的方法或基于机器学习方案模板执行机器学习过程的方法中所描述的某些操作可通过软件方式来实现,某些操作可通过硬件方式来实现,此外,还可通过软硬件结合的方式来实现这些操作。

[0216] 处理器可运行存储在存储部件之一中的指令或代码,其中,所述存储部件还可以存储数据。指令和数据还可经由网络接口装置而通过网络被发送和接收,其中,所述网络接口装置可采用任何已知的传输协议。

[0217] 存储部件可与处理器集成为一体,例如,将RAM或闪存布置在集成电路微处理器等之内。此外,存储部件可包括独立的装置,诸如,外部盘驱动、存储阵列或任何数据库系统可使用的其他存储装置。存储部件和处理器可在操作上进行耦合,或者可例如通过I/O端口、

网络连接等互相通信,使得处理器能够读取存储在存储部件中的文件。

[0218] 此外,所述计算装置还可包括视频显示器(诸如,液晶显示器)和用户交互接口(诸如,键盘、鼠标、触摸输入装置等)。计算装置的所有组件可经由总线和/或网络而彼此连接。

[0219] 根据本发明示例性实施例的用于创建机器学习方案模板的方法或基于机器学习方案模板执行机器学习过程的方法所涉及的操作可被描述为各种互联或耦合的功能块或功能示图。然而,这些功能块或功能示图可被均等地集成为单个的逻辑装置或按照非确切的边界进行操作。

[0220] 例如,如上所述,根据本发明示例性实施例的用于创建机器学习方案模板的装置或基于机器学习方案模板执行机器学习过程的装置可包括存储部件和处理器,其中,存储部件中存储有计算机可执行指令集合,当所述计算机可执行指令集合被所述处理器执行时,执行上文述及的用于创建机器学习方案模板的方法或基于机器学习方案模板执行机器学习过程的方法。

[0221] 以上描述了本发明的各示例性实施例,应理解,上述描述仅是示例性的,并非穷尽性的,本发明不限于所披露的各示例性实施例。在不偏离本发明的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。因此,本发明的保护范围应该以权利要求的范围为准。

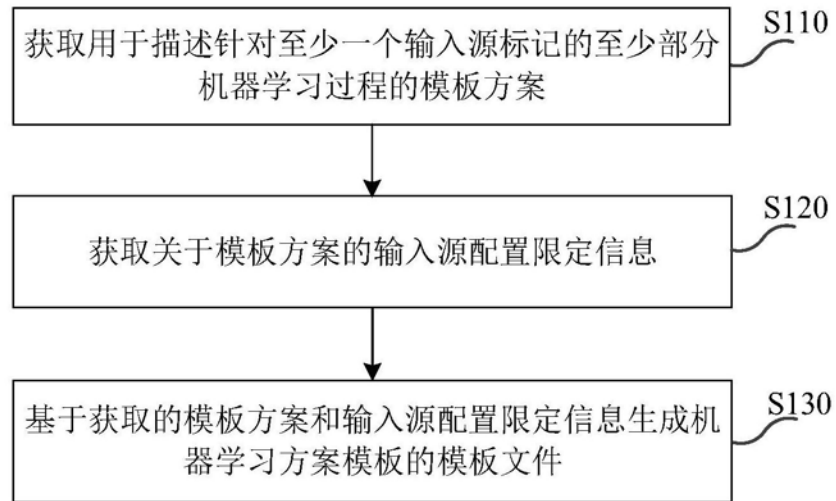


图1

建模模板名称：

建模模板可见状态：

对所有用户可见

对所有用户不可见

建模模板DAG：

点击上传

建模模板配置说明文档（该文档向用户展示，说明配置字段的业务含义）：

点击上传

建模模板输入配置

添加输入表

输入表1名称：

输入表1

nodeID：

添加字段

输入表1字段名

是否需要字段校验

字段类型

字段名1

是

否

Int

输入表2名称：

输入表1

nodeID：

添加字段

输入表1字段名

是否需要字段校验

字段类型

字段名1

是

否

Int

建模模板DAG变量配置

添加变量

\$input1:

脚本替换变量

单选选择项

选项名

选择项1展示信息（默认项）：

选择项2展示信息：

添加选项

选择项字段类型：

Int

选择项1实际信息：

选择项2实际信息：

添加分类线

\$input2:

脚本替换变量

输入项

参数显示名：

参数默认项：

是否校验字段类型

字段类型 

Int

是否校验字段阈值

正则表达式：

正则表达式提示文案：

上分类线名称：

\$input3\$:

运行参数

Node别名：

参数显示名：

参数名：

参数默认项：

是否校验字段类型

字段类型 

int

是否校验字段阈值

正则表达式：

正则表达式提示文案：

添加分类线

保存方案

取消

图2

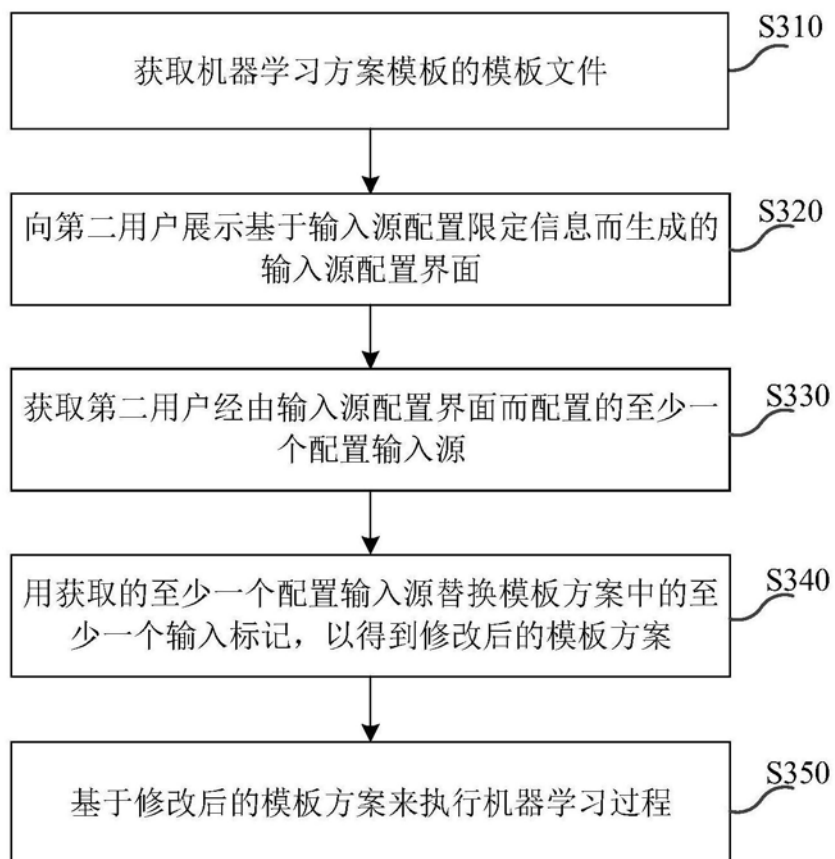


图3

通用营销建模模板配置 ×

建模模板输入配置

建模模板配置说明

用户表: t

user\_id对应: namespace/tablename.后缀

user\_lv对应: modelID/usertable.table

商品表:

sku\_id对应:

a3对应:

age对应字段:

sex对应字段:

user\_reg\_tm对应字段:

a1对应字段:

a2对应字段:

date对应字段:

brand对应字段:

行为表: 请输入表名

user\_id对应字段:

sku\_id对应字段:

data对应字段:

model\_id对应字段:

type对应字段:

时序参数设置

营销时序窗口配置: ☐ 3天 ☐ 5天

理财购买时序窗口配置: ☐ 3天 ☐ 5天

样本参数设置

label配置:

创建DAG进行模型训练

图4

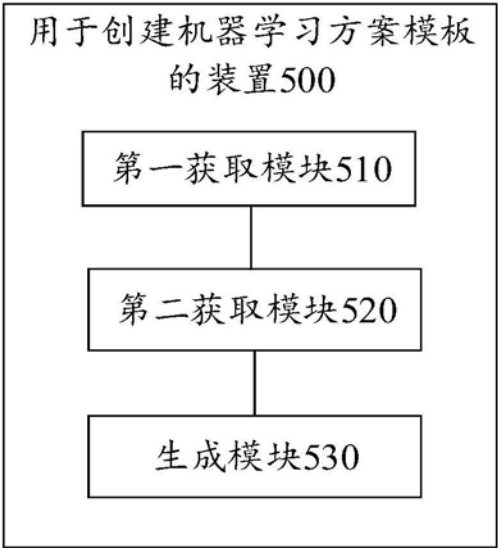


图5

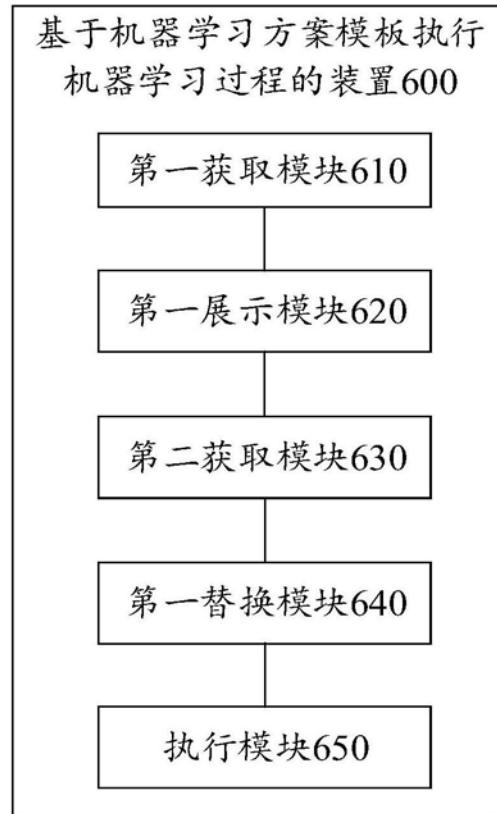


图6