

数据处理模型构建方法、装置、服务器和用户端

技术领域

[01]本说明书属于互联网技术领域，尤其涉及一种数据处理模型构建方法、装置、服务器和用户端。

5 背景技术

[02]随着人工智能技术的发展，通过机器学习构建相应的数据处理模型，再利用所构建的数据处理模型对目标数据进行具体的数据处理，这种数据处理方法在越来越多的领域得到普及应用。

[03]目前，基于现有方法构建上述数据处理模型时，用户需要同时设计并通过对应的代码程序等对多个可能的模型分析进行测试分析等操作（例如包括特征工程、模型选择、模型参数搜索等），以确定出满足目标数据处理要求、合适的模型作为基础模型，再基于该基础模型训练构建所需要的目标数据处理模型。上述方法在实现过程中，对用户的专业能力（例如对模型结构的测试、模型的数据分析等）要求相对较高，并且每一次构建数据处理模型都需要用户手动进行多次测试来找到合适的基础模型，势必会增加处理过程的复杂度，影响处理效率。因此，亟需一种更加简便、高效的数据处理模型构建方法，使得用户不用再繁琐地做测试、分析来寻找合适的模型作为基础模型，从而降低了用户的操作难度，提高了处理效率。

发明内容

[04]本说明书目的在于提供一种数据处理模型构建方法、装置、服务器和用户端，使得用户不用自己繁琐地去做测试、分析来寻找合适的基础模型，构建目标数据处理模型，降低了用户的操作难度，提高了处理效率。

[05]本说明书提供的一种数据处理模型构建方法、装置、服务器和用户端是这样实现的：

[06]一种数据处理模型构建方法，包括：获取目标数据处理模型的模型描述参数和样本数据；根据所述模型描述参数和所述样本数据，确定基础模型；根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。

[07]一种数据处理方法，包括：获取待处理的目标数据；调用目标数据处理模型对所述

目标数据进行数据处理, 其中, 所述目标数据处理模型是根据所获取的目标数据处理模型的模型描述参数和样本数据, 确定基础模型, 再根据所述样本数据对所述基础模型进行训练得到的。

5 [08]一种数据处理方法, 包括: 显示数据输入界面; 接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数, 以及样本数据, 其中, 所述模型描述参数和所述样本数据用于确定基础模型, 所述基础模型用于建立目标数据处理模型。

[09]一种数据处理模型构建装置, 包括: 获取模块, 用于获取目标数据处理模型的模型描述参数和样本数据; 确定模块, 用于根据所述模型描述参数和所述样本数据, 确定基础模型; 训练模块, 用于根据所述样本数据对所述基础模型进行训练, 得到目标数据处理模型。
10

[10]一种服务器, 包括处理器以及用于存储处理器可执行指令的存储器, 所述处理器执行所述指令时实现获取目标数据处理模型的模型描述参数和样本数据; 根据所述模型描述参数和所述样本数据, 确定基础模型; 根据所述样本数据对所述基础模型进行训练, 得到目标数据处理模型。

15 [11]一种用户端, 包括处理器以及用于存储处理器可执行指令的存储器, 所述处理器执行所述指令时实现显示数据输入界面; 接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数, 以及样本数据, 其中, 所述模型描述参数和所述样本数据用于确定基础模型, 所述基础模型用于建立目标数据处理模型。

[12]一种计算机可读存储介质, 其上存储有计算机指令, 所述指令被执行时实现获取目标数据处理模型的模型描述参数和样本数据; 根据所述模型描述参数和所述样本数据, 确定基础模型; 根据所述样本数据对所述基础模型进行训练, 得到目标数据处理模型。
20

[13]本说明书提供的一种数据处理模型构建方法、装置、服务器和用户端, 由于通过有针对性地获取模型的描述参数, 并基于模型的描述参数, 服务器能够准确地确定出用户的建模要求, 并根据用户的建模要求自动匹配到合适的模型作为基础模型, 再进行相应的数据处理模型的构建, 从而使得用户不用再繁琐地去做测试、分析来确定合适的基础模型, 降低了用户的操作难度, 提高了处理效率。
25

附图说明

[14]为了更清楚地说明本说明书实施例或现有技术中的技术方案, 下面将对实施例或现

有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

5 [15]图 1 是基于本说明书实施例提供的数据处理模型构建方法的系统结构的一种实施例的示意图；

[16]图 2 是在一个场景示例中，应用本说明书实施例提供的数据处理模型构建方法所展示的数据输入界面的一种实施例示意图；

[17]图 3 是在一个场景示例中，应用本说明书实施例提供的数据处理模型构建方法筛选候选模型的一种实施例示意图；

10 [18]图 4 是在一个场景示例中，应用本说明书实施例提供的数据处理模型构建方法筛选候选模型的另一种实施例示意图；

[19]图 5 是本说明书实施例提供的数据处理模型构建方法的流程的一种实施例的示意图；

[20]图 6 是本说明书实施例提供的数据处理方法的流程的一种实施例的示意图；

[21]图 7 是本说明书实施例提供的数据处理方法的流程的另一种实施例的示意图；

15 [22]图 8 是本说明书实施例提供的服务器的结构的一种实施例的示意图；

[23]图 9 是本说明书实施例提供的用户端的结构的一种实施例的示意图；

[24]图 10 是本说明书实施例提供的数据处理模型构建装置的结构的一种实施例的示意图。

具体实施方式

20 [25]为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本说明书一部分实施例，而不是全部的实施例。基于本说明书中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都应当属于本说明书保护的范围。

25 [26]考虑到基于现有的数据处理模型构建方法，用户在通过机器学习构建目标数据处理模型的过程中，往往需要自己通过编代码等方式去测试、分析（例如具体可以包括：特征工程，模型选择，模型参数搜索等步骤），从多种可能的模型中确定适合的模型作为

基础模型（或称初始模型）。但上述确定合适的模型作为基础模型对用户的专业性要求相对较高，操作难度相对较大。例如，可能需要用户具有较高的数据分析能力和编程知识，以便对基于各个测试模型所反馈的底层数据（例如程序代码等）进行解读、分析，以确定各个测试模型对所处理的目标数据的适用性程度。此外，由于需要用户自己去寻找合适的基础模型，实施过程相对较为复杂，处理效率也相对较低。

[27]针对产生上述问题的根本原因，本说明书考虑可以将模型的描述，以及相关的基础模型的确定分离出来。具体的，在面向用户一侧，可以允许用户提供较为简便的模型描述参数来表征用户对所需要的基础模型的要求限定，进而平台服务器可以根据模型描述参数自动为用户匹配到合适的模型作为基础模型，以使用户可以通过对该基础模型的训练来构建出自己所需要的数据处理模型，从而使得用户不用在繁琐地去做测试、分析来确定合适的基础模型，降低了用户的操作难度，提高了处理效率。

[28]基于上述想法，本说明书实施方式提供一种数据处理模型构建方法，所述数据处理模型构建方法可以应用于包括服务器和用户端的系统中。可以参阅图 1 所示，其中，上述用户端和服务器相互耦合以进行具体的数据交互。

[29]基于上述系统，用户可以通过用户端设定满足自己建模要求的模型描述参数，以及输入相应的样本数据。服务器可以接收用户端发送的模型描述参数和样本数据，并根据用户设定的模型描述参数和样本数据选择合适的模型作为基础模型以及对应的模型参数，进而可以将该基础模型和模型参数提供给用户，以使用户可以直接利用该基础模型进行目标数据处理模型的训练构建；也可以基于该基础模型和模型参数利用用户输入的样本数据自动进行目标数据处理模型的训练和构建。

[30]在本实施方式中，所述服务器具体可以为具备数据分析、处理，以及数据通信功能的后台业务服务器。所述服务器可以为一个具有数据运算、存储功能以及网络交互功能的电子设备；也可以为运行于该电子设备中，为数据处理、存储和网络交互提供支持的软件。在本实施方式中并不具体限定所述服务器的数量。所述服务器可以为一个服务器，还可以为几个服务器，或者，若干服务器形成的服务器集群。

[31]在本实施方式中，所述用户端可以为具备数据输入、数据展示，以及数据通信功能的端口设备。具体地，所述用户端例如可以是台式电脑、平板电脑、笔记本电脑、智能手机、数字助理、智能可穿戴设备、导购终端、具有网络访问功能的电视机等。或者，所述用户端也可以为能够运行于上述电子设备中的软件应用、图形界面等。

[32]在一个具体的场景示例中，用户可以应用本说明书实施例提供的数据处理模型构建方法，简单、高效地进行目标数据处理模型的训练构建。

[33]在本场景示例中，用户想要构建一种能够识别图片中人物性别的数据处理模型。如果直接基于现有的数据处理模型构建方法，用户需要自己去对多种可能的模型进行测试、分析，以挑选适宜于识别图片人物性别的模型作为基础模型，再进行具体的数据处理模型的训练构建。对用户而言实现较为繁琐，操作难度较大，处理效率也相对较低。

[34]而基于本说明书实施例提供的数据处理模型构建方法，平台服务器可以基于预先定义好的模型描述规则，为用户生成一种可以输入模型描述参数的数据输入界面，并将该数据输入界面发送至用户端。用户端可以向用户展示上述数据输入界面。具体可以参阅图 2 所示。数据输入界面上展示有多个待用户设定的模型描述参数。

[35]其中，上述模型描述参数具体可以理解为一种由用户设定，能够有效地表征出待构建的目标数据处理模型的数据处理要求和/或用户定制要求的参数数据。这样，即使用户对于具体的底层数据不够了解，或者专业能力相对较弱，也可以根据想要构建的目标数据处理模型所要实现的数据处理，简便、高效地通过模型描述参数设置自己的建模要求。并且，上述参数数据符合模型描述规则。即服务器可以基于模型描述规则，通过用户端向用户展示相应的数据输入界面，以使用户能够按照与模型描述规则对应的格式设定具体的模型描述参数，进而可以准确地识别并确定出上述模型描述参数所包含的具体信息，并以上述信息作为依据为用户自动寻找匹配到满足待构建的目标数据处理模型实现要求和/或用户定制要求的模型（即合适的模型），作为基础模型。

[36]具体的，上述模型描述参数可以包括：目标数据处理模型的名称（可以记为 model_name）、数据标签（可以记为 label）、数据特征（可以记为 feature）、评估指标（可以记为 eval_metrics）等。当然，需要说明的是，上述所列举的模型描述参数只是为了更好地说明本说明书实施方式。具体实施时，根据用户的具体情况和待构建的目标数据处理模型的具体特点，还可以引入其他类型的参数数据，例如模型类型等作为模型描述参数。对此，本说明书不作限定。

[37]其中，上述目标数据处理模型的名称可以理解为用户想要构建的数据处理模型的标识符号，例如 A 模型，或者图片人物性别识别模型等。

[38]上述数据标签可以理解为目标数据处理模型对目标数据进行数据处理得到的处理结果。例如，可以是通过对数据处理模型预测或识别得到的结果数据（对应数据处理模型中

的 y 列的数据)。具体的,例如,可以通过数据处理模型识别图片人物性别得到的结果“男”或者“女”。

[39]上述数据特征可以理解为目标数据处理模型在对目标数据进行数据处理的过程中用以作为处理依据的数据特征。例如,可以是数据处理模型从样本数据中提取出来的用于得到预测或识别结果的数据特征(对应数据处理模型中的 x 列的数据)。具体的,例如,可以是图片中人脸部位位置处的毛发浓密程度、图片中人脖子位置处的凸起幅度、图片中人脸位置处的颜色值等。

[40]上述评估指标(也可以称为评估参量)可以理解作为一种能够评价目标数据处理模型数据处理精度、处理效率等性能的参数。

10 [41]具体的,对于不同类型的数据处理模型,对应的评估指标可以不同。对于分类模型,上述评估指标可以包括 AUC (Area Under roc Curve, 曲线下面积) 参数,该指标是一种用于度量模型的分类效果的评估指标,通常分类效果越好,精度越高,对应的 AUC 指标数值越大。此外,对于分类模型,上述评估指标还可以包括:准确率、平均准确率、对数损失函数、精确率-召回率、混淆矩阵等。对于回归模型,上述评估指标可以包括:

15 平方根误差、Quantiles of Errors (误差分位数)、Almost Correct Predictions (近似准确度预测)等。当然,需要说明的是,上述所列举的评估指标只是为了更好地说明本说明书实施方式。具体实施时,可以根据目标数据处理模型所要应用的数据处理的具体特点和用户的具体要求,选择其他合适的性能参数作为上述评估指标。对此,本说明书不作限定。

20 [42]在本场景示例中,用户可以在用户所展示的数据输入界面中的数据处理模型的名称栏中设定名称为“WM”。在数据标签栏中分别设定数据标签 1 为“man”(对应男)、数据标签 2 为“woman”(对应女)。在数据特征栏中分别设定数据特征 1 为图片中人脸的颜色值,数据特征 2 为图片中人脸的光滑程度值,数据特征 3 为图片中人脖子位置的凸起幅度等。在评估指标栏中设定评估指标为“AUC”。

25 [43]此外,上述数据输入界面上还提供有样本数据输入接口,用户可以通过该输入接口导入对应的样本数据。在本场景示例中,上述样本数据具体可以是包含有人物的图片数据。

[44]用户端通过上述数据输入界面接收用户设定的模型描述参数和样本数据,并通过网络将上述数据发送至平台服务器。

[45]平台服务器在接收到模型描述参数和样本数据后,可以根据模型描述参数,结合样

本数据自动地为用户进行测试和分析，进而为用户确定出满足用户要求的模型作为基础模型（也可以称初始模型）；进而可以利用样本数据对该基础模型进行训练以构建得到用户所需要的能够识别图片中人物性别的目标数据处理模型。

- [46]具体实施时，服务器在收到模型描述参数后，可以先对模型描述参数进行解析，以
- 5 确定出用户设定的目标数据处理模型的名称、数据标签、数据特征、评估指标等参数的具体数据。例如，可以通过解析确定出用户设定的目标数据处理模型的名称为“WM”；有两个数据标签，且数据标签 1 为“man”，数据标签 2 为“woman”；有三个数据特征，且数据特征 1 为图片中人脸的颜色值，数据特征 2 为图片中人脸的光滑程度值，数据特征 3 为图片中人脖子位置的凸起幅度；有一个评估指标为 AUC 参数。
- 10 [47]进而，服务器通过对上述数据标签进行分析，可以发现数据标签的数量为 2 而不是 1，且目标数据处理模型所要确定的是目标数据对应的是“man”还是“woman”。因此，可以确定出目标数据处理模型所要执行的数据处理是一种分类类型的数据处理。继而可以从预设的模型类型中进行检索，确定出与上述分类类型的数据处理匹配的模型为分类类型，因此将目标数据处理模型的基础模型的模型类型确定为分类类型。
- 15 [48]在确定基础模型的模型类型为分类类型后，服务器可以检索预设的模型数据库，从满足基础模型的模型类型（即分类模型）的模型中筛选出多个模型作为候选模型，以便后续可以从上述多个候选模型中确定效果较好的模型作为基础模型。
- [49]在本场景示例中，为了能够尽可能从满足基础模型的模型类型的许多模型筛选出可能性较高、效果较好的模型作为候选模型进行后续的测试分析，排除干扰模型，避免对
- 20 一些效果较差的模型浪费时间和资源进行测试分析，可以对所获取的多个满足基础模型的模型类型的模型进行进一步的筛选。
- [50]具体的，为了能够更加准确、快速找到效果较好、较合适的候选模型，可以参阅图 3 所示，将样本数据作为输入，输入至预设的数据处理模型库中已经训练好的，且所应用的基础模型满足上述基础模型的模型类型的数据处理模型（例如数据处理模型 AC、
- 25 数据处理模型 YU、数据处理模型 B2）中，得到各个数据处理模型的数据处理结果（例如预测结果）；再根据各个数据处理模型的数据处理结果统计得到各个数据处理模型的评估指标；再根据目标数据处理模型的评估指标，挑选出与目标数据处理模型的评估指标近似或者一致的数据处理模型 AC 作为推荐模型；并将推荐模型 AC 所应用的基础模型 a_c 确定为候选模型。

[51]此外，还可以根据样本数据的数据类型，对满足基础模型的模型类型的多个模型进行筛选。可以参阅图 4 所示，可以先确定样本数据的数据类型，为图片数据。进而可以确定出目标数据处理模型后续要处理的目标数据也为图片数据。而对模型而言，不同种类的模型有的对图片数据的分析处理准确度比较高，有的对文本数据的分析处理比较高。

- 5 为了提高后续构建的数据处理模型的准确度可以从满足所述基础模型的模型类型的多个模型中筛选出与图片数据匹配(或者适合处理图片数据)的模型作为候选模型。例如，可以提取出模型 a_c、a_r、b_2、m 等模型作为候选模型。

- [52]当然，需要说明的是，上述所列举的对满足基础模型的模型类型中的多个模型进行进一步筛选，以获取效果较好的模型作为候选模型的实现方式只是一种示意性说明，不
10 应构成对本说明书的不当限定。

[53]服务器在确定了多个候选模型后，可以对为多个候选模型中的各个候选模型分别建立预训练模型，再利用上述预训练模型进行测试分析，以选择出效果较好、适用于目标数据处理的，且满足用户要求的模型作为目标数据处理模型的基础模型。

- [54]具体的，服务器可以先获取的样本数据中提取预设比例的样本数据作为测试数据；
15 再对所述测试数据进行预处理，得到处理后的测试数据；进而可以根据所述数据特征，从所述处理后的测试数据中选择出特征数据；根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练，得到所述预训练模型。

- [55]在本场景示例中，具体实施时，服务器可以随机提取样本数据中 10% 的样本数据作为测试数据。当然该数值可以根据具体情况设定，例如，如果想要提高所确定的基础模
20 型的准确度，可以增加该比例值，以增加测试数据的数量。如果想要提高确定基础模型的速度，可以减少该比例值，以减少测试数据的数量。对此，本说明书不作限定。

[56]上述预处理具体可以包括降维处理、清洗处理等等。当然上述所列举的预处理方式只是为了更好地说明本说明书实施方式。具体实施时，可以根据具体的应用场景和要求选择合适的预处理方式。对此，本说明书不作限定。

- 25 [57]在本场景示例中，具体实施时，服务器可以根据数据特征 1 图片中人脸的颜色值，有针对性地提取出测试数据中图片中表征人脸的颜色值的那部分数据作为第一特征数据。类似的，可以分别提取出测试数据中图片中表征人脸的光滑程度值、人脖子位置的凸起幅度的那部分数据作为第二特征数据、第三特征数据。进而可以根据上述第一特征数据、第二特征数据、第三特征数据分别对多个候选模型进行训练，以得到较为简单的

多个预训练模型。其中，每一个预训练模型是基于一种候选模型得到的。

[58]在得到多个预训练模型后，服务器还可以根据预训练模型，通过模型参数搜索确定出各个预训练模型所对应的模型参数。其中，上述模型参数可以理解为预训练模型运行时所要涉及到的相关运行参数。具体的，上述模型参数可以包括：迭代次数、收敛条件、
5 运行时间等等。

[59]进一步，服务器可以对各个预训练模型进行测试，具体的，可以根据评估指标，确定预训练模型和模型参数的适用度；再根据各个预训练模型的适用度，从所述多个候选模型中确定出所述基础模型。从而可以从多个候选模型中筛选出适用于目标数据处理场景、准确度较高，或者满足用户要求的候选模型作为基础模型。

10 [60]其中，所述适用度具体可以理解为用于表征预训练模型和模型参数在针对样本数据的数据处理中的适用性程度。通常对于预训练模型而言，适用度值越高，该预训练模型在进行目标数据的数据处理时准确度也越高，效果也越好。相应的，该预训练模型对应的候选模型在处理该目标数据的数据处理时效果也更为理想。

[61]在本场景示例中，具体实施时，可以随机抽取部分测试数据，对所建立的预训练模型进行测试，得到测试结果。再根据测试结果统计各个预训练模型的 AUC 参数，并根据 AUC 参数确定各个预训练模型的适用度值。例如，如果某个预训练模型的 AUC 参数较高，则该预训练模型的精确度较高，在目标数据的数据处理中的适用性也相对较好，相应得到适用度也相对极高。再根据各个适用度对多个预训练模型进行排序，选择适用度值最高的预训练模型所对应的候选模型（即合适的模型）作为基础模型。当然，也可以
15 以选择适用度值靠前的预设个数的预训练模型对应的多个候选模型或者适用度值大于预设适用度阈值的预训练模型对应的候选模型进行组合，得到组合模型，将该组合模型作为基础模型。

[62]例如，可以将预训练模型 1、2、3、4、5、6 分别对随机抽取的同样的多个图片数据进行分类识别，得到各个预训练模型的对各个图片数据的识别结果，作为各个预训练模型的测试结果。再根据上述测试结果分别统计各个预训练模型的 AUC 参数。进一步根据 AUC 参数确定各个预训练模型的适用度值。比较各个预训练模型的适用度值，发现预训练模型 1 的适用度值最高，因此，可以将预训练模型 1 对应的候选模型确定为目标
25 数据处理模型的基础模型。

[63]在本场景示例中，还可以按照以下方式确定各个预训练模型的适用度，以评价对应

的候选模型的效果：根据所述预训练模型和模型参数，进行参数搜索和结构搜索，得到搜索结果；根据所述搜索结果和所述评估指标，确定所述预训练模型和模型参数的适用度。其中，上述根据所述预训练模型和对应的模型参数进行参数搜索和结构搜索，具体实施时，可以包括：通过以下算法中的至少之一进行参数搜索和结构搜索：蒙特卡洛算法、网格算法、贝叶斯算法、遗传算法等。

[64]当然，需要说明的是上述所列举的确定预训练模型的适用度的方法只是为了更好地说明本说明书实施方式。具体实施，可以根据具体情况选择合适的方式确定预训练模型的适用度。对此，本说明书不作限定。

[65]按照上述方式，服务器可以根据用户简单设定的模型描述参数，准确地匹配到适用于用户待构建的目标数据处理模型的基础模型，并将该基础模型反馈给用户。进而可以利用样本数据对该基础模型进行训练，以构建得到满足用户要求的，能够高效、准确地识别图片中人物性别的 WM 模型，从而高效地完成了目标数据处理模型的构建。

[66]由上述场景示例可见，本说明书提供的数据处理模型构建方法，由于通过有针对性地获取模型的描述参数，并基于模型的描述参数，服务器能够准确地确定出用户的建模要求，并根据用户的建模要求自动匹配到合适的模型作为基础模型，再进行相应的数据处理模型的构建，从而使得用户不用再繁琐地去做测试、分析来确定合适的基础模型，降低了用户的操作难度，提高了处理效率。

[67]参阅图 5 所示，本说明书实施例提供了一种数据处理模型构建方法，其中，该方法具体应用于服务器一侧。具体实施时，该方法可以包括以下内容：

[68]S51：获取目标数据处理模型的模型描述参数和样本数据。

[69]在本实施方式中，上述目标数据处理模型具体可以为待构建的，用于进行目标数据处理的机器学习模型。例如，上述目标数据处理模型可以是用于识别文本数据语义识别模型，也可以是用于区分图片数据中不同颜色的分类模型等等。对于目标数据处理模型的类型和应用本说明书不作限定。

[70]在本实施方式中，上述模型描述参数具体可以为一种由用户设定，能够有效地表征出待构建的目标数据处理模型的数据处理要求和/或用户定制要求的参数数据。

[71]在一个实施方式中，上述模型描述参数具体可以包括以下至少之一：目标数据处理模型的名称、数据标签、数据特征、评估指标等。当然，需要说明的是，上述所列举的模型描述参数只是为了更好地说明本说明书实施方式。具体实施时，根据具体的应用场

景和用户要求还可以引入其他类型的参数作为上述模型描述参数。对此，本说明书不作限定。

[72]其中，上述目标数据处理模型的名称可以理解为为用户想要构建的数据处理模型的标识符号，例如 A 模型，或者图片人物性别识别模型等。上述数据标签可以理解为目标数据
5 数据处理模型对目标数据进行数据处理得到的处理结果。例如，可以是通过对数据处理模型预测或识别得到的结果数据（对应数据处理模型中的 y 列的数据）。具体的，例如，可以是通过对数据处理模型识别图片人物性别得到的结果“男”或者“女”。上述数据特征可以理解为目标数据处理模型在对目标数据进行数据处理的过程中用以作为处理依据的数据特征。例如，可以是数据处理模型从样本数据中提取出来的用于得到预测或识别
10 结果的数据特征（对应数据处理模型中的 x 列的数据）。具体的，例如，可以是图片中人脸部位置处的毛发浓密程度、图片中脖子位置处的凸起幅度、图片中人脸位置处的颜色值等。上述评估指标（也可以称为评估参量）可以理解作为一种能够评价目标数据处理模型数据处理精度、处理效率等性能的参数。例如，可以是 AUC 参数、准确率、平均准确率、对数损失函数、精确率-召回率、混淆矩阵、平方根误差、Quantiles of Errors
15 （误差分位数）、Almost Correct Predictions（近似准确预测值）等。

[73]在本实施方式中，上述样本数据具体可以理解为待处理的目标数据数据类型相同的数据。其中，上述样本数据具体可以是图片数据，也可以是文本数据等等。对于样本数据的数据类型，本说明书不作限定。

[74]S53：根据所述模型描述参数和所述样本数据，确定基础模型。

20 [75]在本实施方式中，上述基础模型具体可以称为用于构建目标数据处理模型的初始模型，具体的上述基础模型可以是具有一定数据处理、分析能力的模型结构（或算法结构），通常需要通过基础模型的训练学习，来构建得到符合用户要求的，能够进行目标数据处理的目标数据处理模型。

25 [76]在本实施方式中，需要说明的是，不同的模型结构在进行具体的数据处理时差异较大。例如，有的模型结构适用于处理文本数据，有的模型结构适用于处理图片数据。或者，有的模型结构适用于解决分类问题，有的模型结构适用于解决回归问题。因此，确定一种适用于目标数据处理的，满足用户要求的基础模型对于后续构建得到目标数据处理模型的使用效果具有较为重要的影响。

[77]在本实施方式中，为了避免用户繁琐地自己去做多次测试分析来找到合适的模型作

为基础模型，服务器可以以用户设定的模型描述参数作为处理依据，结合用户输入的样本数据自动地去做测试分析，来寻找能够进行目标数据处理，满足用户要求的模型（即合适的模型）作为数据处理模型。

[78]在本实施方式中，具体实施时，可以根据模型描述参数确定出用户所要构建的目标数据处理模型的模型特征；再根据模型特征逐步地缩小范围从多个模型结构中筛选得到合适的模型作为基础模型，提供给用户来构建具体的目标数据处理模型。

[79]在一个实施方式中，上述根据所述模型描述参数和所述样本数据，确定基础模型，具体实施时，可以包括：根据所述数据标签，从预设的模型类型组中确定基础模型的模型类型，其中，所述预设的模型类型组中至少包括：分类类型和回归预测类型；从满足所述基础模型的模型类型的模型中，选择出多个模型作为候选模型；根据所述样本数据和所述数据特征，为多个候选模型中的各个候选模型建立预训练模型，并确定各个预训练模型的模型参数；根据所述评估指标，确定各个预训练模型的适用度，其中，所述适用度用于表征基于预训练模型和该预训练模型对应的模型参数对所述样本数据进行数据处理的适用性程度；根据各个预训练模型的适用度，从所述多个候选模型中确定出所述基础模型。

[80]S55：根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。

[81]在本实施方式中，在确定了基础模型后，进一步可以将基础模型作为初始模型；再利用样本数据对所述基础模型进行训练，以构建得到符合要求的目标数据处理模型。当然，需要说明说明，上述所列举的利用基础模型进行训练以构建得到目标数据处理模型只是应用基础模型的一种实现方式。具体实施时，根据具体的应用场景和技术的发展程度还可以利用所确定的基础模型进行其他方面的应用。对此，本说明书不作限定。

[82]在本实施方式中，通过将模型描述、确定基础模型，与数据处理模型的构建分离，用户仅需提供模型描述参数，服务器即可以根据用户提供的模型描述参数自动进行模型描述，并自动确定出合适的基础模型提供给用户，以便可以基于上述基础模型再进行数据处理模型的构建。从而使得用户不需要自己再繁琐地寻找合适的基础模型，减轻了用户的负担。

[83]由上可见，本说明书提供的数据处理模型构建方法，由于通过有针对性地获取模型的描述参数，并基于模型的描述参数，服务器能够准确地确定出用户的建模要求，并根据用户的建模要求自动匹配到合适的模型作为基础模型，再进行相应的数据处理模型的

构建,从而使得用户不用再繁琐地去做测试、分析来确定合适的基础模型,降低了用户的操作难度,提高了处理效率。

[84]在一个实施方式中,所述模型描述参数具体可以包括以下至少之一:目标数据处理模型的名称、数据标签、数据特征、评估指标等。当然,需要说明的是,上述所列举的模型描述参数只是一种示意性说明。具体实施时,还可以根据具体情况引入其他类型的参数作为上述模型描述参数。对此,本说明书不作限定。

[85]在一个实施方式中,所述模型描述参数具体还可以包括:基础模型的模型类型、样本数据的拆分规则、阈值参数等等。对于具有一定专业能力的用户,例如工程师,可以为他们提供更多的模型参数的设定权限,以便获取更多的模型参数,从而可以基于更多地模型参数更加快速、精确地找到合适的模型作为目标数据处理模型的基础模型。

[86]在一个实施方式中,根据所述模型描述参数和所述样本数据,确定基础模型,具体实施时,可以包括以下内容:

[87]S1:根据所述数据标签,从预设的模型类型组中确定基础模型的模型类型,其中,所述预设的模型类型组中至少包括:分类类型和回归预测类型。

[88]在本实施方式中,上述根据所述数据标签,从预设的模型类型组中确定基础模型的模型类型,具体可以理解为分析数据标签数量,以及数据标签的数据和数据标签的相互关系,来确定待构建的目标数据处理模型所要执行的目标数据处理的类型;再根据目标数据处理的类型确定出适用于该目标数据处理的的基础模型的类型。

[89]在本实施方式中,还需要补充的是,通过分析评估指标也可以判断模型类型。因此,具体实施时,除了可以根据数据标签外,还可以根据评估指标,或者数据标签和评估指标的组合,从预设的模型类型组中确定出基础模型的模型类型。

[90]在本实施方式中,上述预设的模型类型组中至少包括有分类类型和回归预测类型等。其中,对于每一种类型而言,分别可以包含有多个具体的模型结构。当然,需要说明的是,上述所列举的预设的模型类型组中所包含的模型类型只是一种示意性说明。根据具体的应用场景和技术的发展情况,上述预设的模型类型组中还可以包含其他的模型类型。对此,本说明书不作限定。

[91]在本实施方式中,需要补充的是,模型类型为分类类型的模型往往相对较适用于解决涉及分类问题的数据处理;分类类型为回归预测类型的模型往往相对较适用于解决涉及预测问题的数据处理。

[92]S2: 从满足所述基础模型的模型类型的模型中, 选择出多个模型作为候选模型。

[93]在本实施方式中, 上述候选模型具体可以理解为适用于目标数据处理的(即满足基础模型的模型类型), 有可能用作目标数据处理模型的基础模型的模型结构。具体的, 后续可以对上述后选模型进行进一步的测试筛选, 以候选模型中确定出基础模型。

- 5 [94]S3: 根据所述样本数据和所述数据特征, 为多个候选模型中的各个候选模型建立预训练模型, 并确定各个预训练模型的模型参数。

[95]在本实施方式中, 上述预训练模型具体可以理解为基于候选模型得到的相对较为简单的数据处理模型。

- 10 [96]在本实施方式中, 上述模型参数具体可以理解为可以理解为预训练模型运行时所要涉及到的相关运行参数。具体的, 上述模型参数可以包括: 迭代次数、收敛条件、运行时间等等。当然, 上述所列举的模型参数只是一种示意性说明。对于模型参数的具体内容, 本说明书不作限定。

- 15 [97]在本实施方式中, 上述根据所述样本数据和所述数据特征, 为多个候选模型中的各个候选模型建立预训练模型, 具体可以包括: 从所述样本数据中提取测试数据; 根据数据特征从测试数据中提取特征数据; 利用上述特征数据对各个候选模型分别进行训练, 以得到与各个候选模型对应的预训练模型。

[98]其中, 上述特征数据可以理解为样本数据中包含有数据特征的部分数据。例如, 上述特征数据可以是表征图片中人脸颜色的数据。

- 20 [99]在本实施方式中, 在得到多个预训练模型后, 进一步, 还可以根据具体的预训练模型, 通过模型参数搜索确定出各个预训练模型所对应的模型参数。

[100] S4: 根据所述评估指标, 确定各个预训练模型的适用度, 其中, 所述适用度用于表征基于预训练模型和该预训练模型对应的模型参数对所述样本数据进行数据处理的适用性程度。

- 25 [101] 在本实施方式中, 上述适用度具体可以理解为用于表征预训练模型和模型参数在针对样本数据的数据处理中的适用性程度。通常对于预训练模型和对应的模型参数而言, 适用度值越高, 该预训练模型基于对应的模型参数在进行目标数据的数据处理时准确度也越高, 效果也越好。相应的, 该预训练模型所对应的候选模型基于上述模型参数在处理该目标数据的数据处理时效果也更为理想。

- [102] 在一个实施方式中, 上述根据所述评估指标, 确定各个预训练模型的适用度, 具体可以包括以下内容: 分别利用各个预训练模型和对应的模型参数对测试数据进行数据处理, 得到数据处理结果作为各个预训练模型和模型参数的测试结果; 根据各个预训练模型和模型参数的测试结果统计各个预训练模型的评估指标; 根据各个预训练模型的
- 5 指标参确定各个预训练模型和模型参数的适用度。需要说明的是, 上述所列举的适用度的确定方式只是一种示意性说明。具体实施时, 还可以根据具体情况和要求, 采用其他合适的方式确定预训练模型的适用度。对此, 本说明书不作限定。
- [103] S5: 根据各个预训练模型的适用度, 从所述多个候选模型中确定出所述基础模型。
- 10 [104] 在本实施方式中, 上述根据各个预训练模型的适用度, 从所述多个候选模型中确定出所述基础模型, 具体实施时, 可以包括: 比较各个预训练模型的适用度, 将适用度最高的预训练模型所对应的模型确定为目标数据处理模型的基础模型。也可以包括: 将适用度大于预设适用度阈值的多个预训练模型所对应的模型进行组合, 得到组合模型; 并将所述组合模型确定为所述基础模型等。当然, 上述所列举的根据适用度从候选模型
- 15 中确定出合适的基础模型只是一种示意性说明。具体实施时, 根据具体情况, 也可以采用其他合适的方式来根据适用度确定出基础模型。对此, 本说明书不作限定。
- [105] 在一个实施方式中, 从满足所述基础模型的模型类型的模型中, 选择出多个模型作为候选模型, 具体实施时, 可以包括以下内容:
- [106] S1: 分别利用预设的数据处理模型库中所应用的基础模型满足所述基础模型的
- 20 模型类型的数据处理模型对所述样本数据进行数据处理, 得到数据处理模型的数据处理结果;
- [107] S2: 根据所述数据处理模型的数据处理结果, 统计数据处理模型的评估指标;
- [108] S3: 将评估指标与目标数据处理模型的评估指标的差异程度小于预设程度阈值的数据处理模型确定为推荐模型;
- 25 [109] S4: 将所述推荐模型所应用的基础模型确定为候选模型。
- [110] 在本实施方式中, 上述预设的数据处理模型库中具体存储有多种不同的已经训练好的数据处理模型。需要补充的是, 上述预设的数据处理模型库可以每隔预设时间段获取最新的训练好的数据处理模型对预设的数据处理模型库进行更新。
- [111] 在本实施方式中, 上述评估指标与目标数据处理模型的评估指标的差异程度小

于预设程度阈值，具体可以理解为数据处理模型的评估指标与目标数据处理模型的评估指标相似或者一致。其中，上述预设程度阈值具体可以根据精度要求灵活设定。对此，本说明书不作限定。

5 [112] 在本实施方式中，上述推荐模型可以理解为所适用的数据处理与待构建的目标数据处理模型所要执行的目标数据处理相似或者一致的，且已经训练好的数据处理模型。这类数据处理模型对于构建目标数据处理模型具有较高的参考借鉴价值。

10 [113] 在本实施方式中，具体实施时，可以将所获取的样本数据作为输入，分别输入至预设的数据处理模型中所应用的基础模型满足上述基础模型的模型类型的数据处理模型中，得到上述各个数据处理模型的数据处理结果；再分别根据各个数据处理模型的数据处理结果，统计得到各个数据处理模型的评估指标；再以目标数据处理模型的评估指标作为依据，找到与目标数据处理模型的评估指标一致或相似的数据处理模型确定为推荐模型，再将推荐模型所应用的模型确定为所述候选模型，从而可以更加快速、高效地找到效果较好的候选模型。

15 [114] 在一个实施方式中，为了进一步缩小候选模型的数量，提高确定基础模型的效率，具体实施时，所述方法还可以包括以下内容：

[115] S1：确定所述样本数据的数据类型，其中，所述样本数据的数据类型包括以下至少之一：图片数据、文本数据；

[116] S2：根据所述样本数据的数据类型，从所述基础模型的模型类型中筛选出与所述样本数据的数据类型匹配的模型类型作为所述基础模型的模型类型。

20 [117] 在本实施方式中，上述样本数据的数据类型具体可以包括图片数据、文本数据、图像数据、数值数据等等。当然，需要说明的是上述所列举的数据类型只是为了更好地说明本说明书实施方式。对于样本数据的数据类型，本说明书不作限定。

25 [118] 在本实施方式中，需要补充的，不同的模型结构往往所适用的数据类型也不相同。例如，有的模型可能适用于处理图像数据，而不适用于处理文本数据。有的模型则可能适用于处理文本数据，而不适用于处理数值数据。因此，为了进一步缩小候选模型的数量，更加快速地确定出适用于目标数据处理的基础模型，可以结合样本数据的数据类型对基础模型的模型类型做进一步的限定。例如，数据处理涉及的解决问题是分类问题，且样本数据的数据类型为文本数据，则可以综合上述情况，确定基础模型的模型类型为分类类型，且适用于处理文本数据（即与文本数据匹配）的模型类型。这样可以进

一步缩小可以作为基础模型的模型类型范围，从而可以减少基于基础模型的模型类型所获取的候选模型的数量，后续只要对更少的候选模型进行测试便可以确定出合适的基础模型。从而达到降低工作量，提高确定效率的目的。

5 [119] 在一个实施方式中，上述根据所述样本数据和所述数据特征，为多个候选模型中的各个候选模型建立预训练模型，具体实施时，可以包括以下内容：

[120] S1：从所述样本数据中提取预设比例的样本数据作为测试数据。

10 [121] 在本实施方式中，具体实施时时可以从样本数据中按照预设比例随机提取部分样本数据作为测试数据。其中，上述预设比例可以根据具体的应用场景和精度要求灵活设定。例如，如果想要提高所确定的基础模型的准确度，可以增加该比例值，以增加测试数据的数量。如果想要提高确定基础模型的速度，可以减少该比例值，以减少测试数据的数量。对此，本说明书不作限定。

[122] S2：对所述测试数据进行预处理，得到处理后的测试数据，其中，所述预处理包括：降维处理和/或清洗处理。

15 [123] 在本实施方式中，上述预处理具体可以包括：降维处理、清洗处理等等。当然上述所列举的预处理方式只是为了更好地说明本说明书实施方式。具体实施时，可以根据具体的应用场景和要求选择合适的预处理方式。例如，样本数据的数据量比较大，所包含的信息比较多，因此，可以通过降维处理对测试数据进行预处理，以便提高后续数据处理的效率。又例如，对于一些随机性较强的数据处理，要求用来训练的样本数据之间尽量不存在一定顺序上的联系，因此，可以先通过清洗数据对测试数据进行预处理，
20 以降低由于测试数据之间的相互联系对模型产生的误差。

[124] S3：根据所述数据特征，从所述处理后的测试数据中选择出特征数据。

[125] 在本实施方式中，根据所述数据特征，从所述处理后的测试数据中选择出特征数据，具体可以包括：从处理后的测试数据中，抽取包含有数据特征的部分数据作为上述特征数据（也可以称为特征工程）。

25 [126] S4：根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练，得到所述预训练模型。

[127] 在本实施方式中，上述预训练模型具体可以理解为基于一个候选模型通过较为简单的训练得到的，具有一定数据处理精度的数据处理模型。

[128] 在本实施方式中，根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练，得到所述预训练模型，具体可以包括：利用所抽取的特征数据分别对各个候选模型进行训练，从而可以得到多个相对较为简单的预训练模型。其中，上述预训练模型分别和一个候选模型对应。

- 5 [129] 在一个实施方式中，上述根据所述评估指标，确定各个预训练模型的适用度，具体实施时，可以包括以下内容：根据所述预训练模型和模型参数，进行参数搜索和结构搜索，得到搜索结果；根据所述搜索结果和所述评估指标，确定所述预训练模型和模型参数的适用度。

- 10 [130] 在本实施方式中，上述根据所述预训练模型和对应的模型参数进行参数搜索和结构搜索，具体可以包括：通过以下算法中的至少之一进行参数搜索和结构搜索：蒙特卡洛算法、网格算法、贝叶斯算法、遗传算法等。当然，需要说明的是，上述所列举的用于进行参数搜索和结构搜索的算法只是一种示意性说明。具体实施时，根据具体的应用场景和精度要求还可以根据其他算法进行参数搜索和结构搜索。对此，本说明书不作限定。

- 15 [131] 在本实施方式中，上述根据所述搜索结果和所述评估指标，确定所述预训练模型和模型参数的适用度，具体实施时，可以包括：根据各个预训练模型和模型参数的搜索结果，以及用户设定的评估指标对各个预训练模型和对应的模型参数进行评分，并将该评分数值作为预训练模型和模型参数的适用度。

- 20 [132] 在一个实施方式中，上述根据各个预训练模型的适用度，从所述多个候选模型中确定出所述基础模型，具体实施时，可以包括：将所述预训练模型中适用度最高的预训练模型所使用的模型确定为所述基础模型。

- 25 [133] 在一个实施方式中，上述根据所述适用度，根据各个预训练模型的适用度，从所述多个候选模型中确定出所述基础模型，具体实施时，还可以包括：将所述预训练模型中适用度大于预设适用度阈值（或者适用度排序前几位）的多个预训练模型所使用的模型进行组合，得到组合模型；并将所述组合模型确定为所述基础模型。这样可以综合多种不同的效果较好的模型结构的优势，得到更加适用于目标数据处理的模型结构作为基础模型。

[134] 在一个实施方式中，所述模型描述参数还可以包括：基础模型的模型类型、样本数据的拆分规则等。其中，上述样本数据的拆分规则具体可以包括有用户设定测试数

据的预设比例,以及拆分样本数据的指定方式等。这样可以为具备不同的专业能力的用户提供更多选择的模型描述参数进行设定,以便能够更好地满足用户的个性化要求,更加准确、快速地确定出基础模型和模型参数。

[135] 由上可见,本说明书提供的数据处理模型构建方法,由于通过有针对性地获取模型的描述参数,并基于模型的描述参数,服务器能够准确地确定出用户的建模要求,并根据用户的建模要求自动匹配到合适的模型作为基础模型,再进行相应的数据处理模型的构建,从而使得用户不用再繁琐地去做测试、分析来确定合适的基础模型,降低了用户的操作难度,提高了处理效率;还具体通过依据数据标签、样本数据的数据类型等多种特征筛选合适的候选模型来确定基础模型,排除了部分干扰,减少了工作量,从而进一步提高了处理效率。

[136] 参阅图 6 所示,本说明书实施例提供了一种数据处理方法,用于进行具体的目标数据处理。具体实施时,该方法可以包括以下内容:

[137] S61: 获取待处理的目标数据。

[138] 在本实施方式中,上述待处理的目标数据具体可以是待分类的图片数据,也可以是待识别的文本数据,还可以是待预测的图像数据等。对于上述待处理的目标数据的数据类型和数据内容,本说明书不作限定。

[139] S63: 调用目标数据处理模型对所述目标数据进行数据处理,其中,所述目标数据处理模型是根据所获取的目标数据处理模型的模型描述参数和样本数据,确定基础模型,再根据所述样本数据对所述基础模型进行训练得到的。

[140] 在本实施方式中,上述目标数据处理模型是服务器预先根据用户设定的模型描述参数和输入的样本数据,匹配得到能够适用于目标数据处理,且满足用户的要求的合适模型作为基础模型;再通过利用样本数据对上述基础模型进行训练、学习所构建得到的数据处理模型。

[141] 在本实施方式中,上述调用目标数据处理模型对所述目标数据进行数据处理,具体可以包括:将目标数据作为输入,输入将上述目标数据处理模型中;目标数据处理模型对该目标数据处理模型进行相应的数据处理,以得到对应的数据处理结果。例如,将包含有人物的图片输入至能够识别图片中人物性别的目标数据处理模型中;目标数据处理模型对图片进行相应的数据处理,得到数据处理结果为:图片中人物为女,从而完成了对目标数据的数据处理。

[142] 参阅图 7 所示, 本说明书实施例还提供了一种数据处理方法, 该方法可以用于用户端, 用于获取目标数据处理模型的模型描述参数。具体实施时, 该方法可以包括以下内容:

[143] S71: 显示数据输入界面。

5 [144] 在本实施方式中, 上述数据输入界面具体可以是如图 2 所示的图像输入界面, 也可以是一种代码输入界面(即提供给用户设置定义的代码模板)等。当然, 需要说明的是, 上述所列举的数据输入界面只是为了更好地说明本说明书实施方式, 对于数据输入界面的具体形式, 本说明书不作限定。

10 [145] 在本实施方式中, 需要补充的是, 通过输入界面输入模型描述参数只是一种示意性说明, 具体实施时, 用户还可以通过命令行、交互式, 例如 presto(一种分布式 SQL 查询引擎)、ipython-notebook(一种基于 Python 的编程工具)等, 或者合并 API 或程序中输入接口等输入目标数据处理模型的模型参数。

15 [146] S73: 接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数, 以及样本数据, 其中, 所述模型描述参数和所述样本数据用于确定基础模型, 所述基础模型用于建立目标数据处理模型。

20 [147] 在本实施方式中, 需要说明的是, 现有的数据处理模型构建方法, 由于将模型描述、基础模型的确定与模型的训练构建混合在一起, 导致实施时缺乏灵活性, 不容易拓展, 对开发者也不够友好, 不能跟踪开发全过程, 没法借助版本管理工具, 例如 git, 进行灵活的帮助控制和代码审核, 并且也不易于更新对某个问题升级新的解法模型。即现有方法对用户操作难度较高, 不够方便、友好。

25 [148] 为了解决上述问题, 可以事先通过语言定义, 定义出模型描述参数, 这样用户可以有针对性地输入或者设置具体的满足目标数据处理以及用户定制化要求的模型描述参数, 机器可以按照一定的语法规则获取用户设定的具体的模型描述参数, 并生成相应的执行文件, 例如代码程序, 进而根据执行文件中用户设定的模型描述参数自动寻找合适的基础模型, 提供给用户; 再对所寻找到的基础模型进行训练, 构建具体的目标数据处理模型, 从而将模型描述、基础模型的确定与模型的训练构建分离, 降低数据处理模型的构建难度和复杂度, 提高用户体验。

[149] 在本实施方式中, 具体实施时, 可以采用 SQL (Structured Query Language, 结构化查询语言)语言按照一定语法规则(例如通过关键词方式)定义上述模型描述参数。

例如，生成关于模型描述参数的用于定义的代码，再基于上述用于定义的代码向用户显示相应的数据输入界面，以接收用户设定的模型描述参数。当然，需要说明的是，上述所列举的定义语言以及语法规则只是为了更好地说明本说明书实施方式。具体实施时，可以根据具体情况选择使用其他语言或者语法规则定义上述模型描述参数。对此，本说明书不作限定。

[150] 在本实施方式中，上述模型描述参数具体可以包括：目标数据处理模型的名称、数据标签、数据特征、评估指标等，对应的关键词分别为：model_name、label、feature、eval_metrics 等。当然，需要说明的是，上述所列举的模型描述参数只是为了更好地说明本说明书实施方式。不应构成对本说明书的不当限定。

10 [151] 在本实施方式中，具体实施时，可以按照以下方式进行具体的定义，得到包含有模型描述参数的用于定义的代码： TRAIN model_name [[BY model_type] [WITH params]] [FOLD number] [PARTITION partition_fields]
SELECT [LABEL label_names]
EVAL eval_metrics [SELECT]
[TIME time_limit]。

[152] 其中，上述用于定义的代码包含有四个部分，分别为训练部分、数据部分、评估部分和计算限制条件部分。

[153] 上述 TRAIN model_name [[BY model_type] [WITH params]] [FOLD number] [PARTITION partition_fields]
用于定义训练部分，其中， []表示可选部分，TRAIN 表示所述训练部分的关键词，model_name 表示训练产生的模型名称；BY 用于指定模型类型 model_type 进行训练；WITH 用于指定模型参数 params；FOLD 用于指定层 k-fold 的取值 number；PARTITION 用于指定用于分区的数据字段 partition_fields；。需要补充的是，上述
的具体含义可以参考相关技术，例如 HTML 5。这里可以是表示为一种常用的描述，例如将问题分成多种子问题，将模型分成子模型。

[154] 上述 SELECT [LABEL label_names]
用于定义数据部分，其中， []表示可选部分，SELECT 表示使用一个或多个数据字段，LABEL 用于指定一个或多个标签 label_names。例如，可以缺省用第一个数据字段作为标签 LABEL，如果 LABEL 多过一个，则可以用 LABEL 指定 label_names
。

[155] 上述 EVAL eval_metrics [SELECT]
用于定义评估部分，其中， []表示可选部分，EVAL 用于指定评估指标 eval_metrics。如果用不同的数据来评估，可以在评估部分再添加评估数据部分
。

[156] 上述[TIME time_limit]用于定义计算限制条件部分,其中,[]表示可选部分,TIME用于指定模型训练完成的时间限制。例如,可以用 TIME 指定模型训练完成的时间限制;如果没有指定,则可以完全交给系统处理。

[157] 进一步,考虑到对所述训练模型进行离线计算,上述用于定义的代码还可以包括离线计算部分,表示为:
SCORE mode_name TO table_name INDEX Index_list (创建 table/insert with index list)
SELECT
。其中,
SCORE 用于选取训练好的模型名称; mode_name TO 用于定义计算完后存储的数据表 table_name; INDEX 用于定义所述数据表的索引字段 Index_list,这些 index 字段也将出现在 SELECT 部分;
SELECT 用于定义用于离线计算的数据。例如可以要求去掉 INDEX 部分后和训练模型的非 LABEL 字段完全一致。

[158] 进一步,考虑到对所述训练模型进行在线计算,上述用于定义的代码还可以包括在线计算部分,表示为:
DEPLOY model_name WITH params
。其中, DEPLOY 用于将所述训练模型 model_name 部署成接口服务(例如 rest api), WITH 用于根据数据中心的要求添加相应的参数配置 params。其中,
语言可以通过嵌套,拓展到多个模型组合的情况。

[159] 当然,上述所列举的用于定义的代码只是一种示意性说明,不应构成对本说明书的不当限定。语言的具体描述可以是演进或变化的。例如还可以描述如下形式:

```
CREATE TABLE model1 AS
SELECT
model_train (features [, labels] [, options]) FROM source_table1;

CREATE TABLE predict1 AS SELECT model_predict (model1 [, options]) FROM
source_table2.
```

[160] 在本实施方式中,对于图像输入界面,参阅图 2 所示,可以通过接收各个数据栏中的数据信息获取目标数据处理模型的模型描述参数和样本数据。例如,可以通过数据输入界面中的数据标签栏获取用户设定的数据标签,通过数据特征栏获取用户设定的数据特征,通过评估指标栏获取用户设定的评估指标,通过目标数据处理模型的名称栏获取用户设定的目标数据处理模型的模型名称,通过样本数据输入接口获取用户输入的样本数据。

[161] 在本实施方式中,对于代码输入界面,可以通过接收用户在代码输入界面中的

预设位置处设置的字符串来获取目标数据处理模型的模型描述参数，通过接收用户触发的预设的代码指令来获取指定位置处的相应的样本数据。

- [162] 具体的，用户可以在上述代码输入界面中的预设位置处定义或者设置具体的字符串作为模型描述参数。例如，用户端可以通过 SELECT 指令接收代码输入界面中
- 5 “TRAIN model_name” 中下划线位置的用户设置的字符串作为目标数据处理模型的模型名称。通过 SELECT 指令接收 “LABEL label_names” 中下划线位置的用户设置的字符串作为目标数据处理模型的数据标签。通过 SELECT 指令接收 “EVAL eval_metrics” 中下划线位置的用户设置的字符串作为目标数据处理模型的评估指标等。对于其他模型描述参数的获取可以参照上述实施方式，本说明书不在赘述。
- 10 [163] 在本实施方式中，用户端在通过上述方式获取了模型描述参数和样本数据后，可以将上述模型描述参数和样本数据以代码等执行文件的形式发送至服务器；服务器可以通过解析（基于 SQL 的解析）代码等执行文件，获取用户设定的模型描述参数；根据用户设定的模型描述参数确定基础模型；再根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。
- 15 [164] 在本实施方式中，服务器在接收到包含有模型描述参数的代码等执行文件时，具体实施时，可以相应的将执行文件分成 4 个部分进行分析处理。
- [165] 第一部分是训练部分，具体可以利用 TRAIN 关键词提取 model_name 作为训练产生的数据处理模型的模型名称；可以利用 BY 关键词指定用哪些 model_type（即模型类型）进行训练；可以利用 WITH 关键词指定相关模型的参数 params，其中，模型参数
- 20 具体可以用 Json 描述的 FOLD 指定 k-fold 的取值；可以利用 PARTITION 关键词指定用于分区的数据字段 partition_fields。
- [166] 第二个部分是数据部分，具体可以利用 SELECT 关键词通过 SQL 语法，缺省用第一个数据字段作为标签 LABEL。具体的，如果 LABEL 多过一个，则用 LABEL 指定 label_names。
- 25 [167] 第三部分是评估部分，具体可以利用 EVAL 关键词指定评估指标 eval_metrics。如果用不同的数据来评估，可以再评估部分添加评估数据部分。
- [168] 第四部分是计算限制性条件部分，具体可以利用 TIME 关键词指定模型训练完成的时间限制。如果用户没设定时间限制，则可以完全交给服务器自行处理。
- [169] 进一步，对于服务器，训练好的模型还可以按照以下方式用于离线计算：CORE

mode_name TO table_name INDEX Index_list
SELECT。其中，模型离线计算具体可以分成两个部分：SCORE 选取训练好的模型名字 mode_name TO 定义计算完存储的数据表 INDEX 定义数据表的 index 字段，其中，这些 index 字段也将出现在 SELECT 部分。而 SELECT 定义用于离线计算的数据，要求去掉 INDEX 部分后和训练 model 的非 label 字段完全一致。对于模型在线计算可以按照以下方式执行：DEPLOY model_name WITH params 将模型 deploy 成一个 rest api 服务，WITH 根据数据中心的要求添加相应的参数配置 params。

- [170] 这样服务器通过解析可以获取得到由用户设定的具体的模型描述参数，以便后续可以利用上述模型描述参数自动确定出合适的基础模型。
- 10 [171] 在一个实施方式中，考虑到对于部分专业能力较强的用户，可以为他们提供更大的自由度和权限来设置模型描述参数，以便服务器能够更加快速、精准地确定出合适的基础模型。因此，在接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数，以及样本数据后，所述方法具体实施时，还可以包括以下内容：响应展示操作，展示基于所述模型描述参数生成的执行文件；接收对所述执行文件的修改操作，并根据
- 15 所述修改操作更新所述执行文件。

- [172] 在本实施方式中，上述执行文件可以理解为具体的程序代码。其中，上述执行文件中包含有模型描述参数，以及基于模型描述参数自动生成其他的模型设置参数。
- [173] 在本实施方式中，上述数据输入界面中具体可以包含有展示键或者展示执行文件的指令。用户可以通过点击上述展示键或者调用展示执行文件的指令来发出展示操作。
- 20 继而用户端在接收到用户的展示操作，展示出相应的执行文件。

- [174] 用户可以对上述执行文件中所包含的模型描述参数，以及其他模型设置参数直接进行修改操作。用户端可以接收用户对执行文件的修改操作，并根据用户的修改操作更新所述执行文件，并将更新后的执行文件发送至服务器，以便服务器可以根据用户修改后的执行文件更加快速、准确地确定基础模型，来构建目标数据处理模型。
- 25 [175] 本说明书实施例还提供一种服务器，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器具体实施时可以根据指令执行以下步骤：获取目标数据处理模型的模型描述参数和样本数据；根据所述模型描述参数和所述样本数据，确定基础模型；根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。

[176] 为了能够更加准确地完成上述指令，参阅图 8 所示，本说明书还提供了另一种

具体的服务器，其中，所述服务器包括网络通信端口 801、处理器 802 以及存储器 803，上述结构通过内部线缆相连，以便各个结构可以进行具体的数据交互。

[177] 其中，所述网络通信端口 801，具体可以用于获取目标数据处理模型的模型描述参数和样本数据。

5 [178] 所述处理器 802，具体可以用于根据所述模型描述参数和所述样本数据，确定基础模型；根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。

[179] 所述存储器 803，具体可以用于存储所获取的模型描述参数、样本数据、处理器 902 生成的中间数据，以及存储相应的指令程序。

10 [180] 在本实施方式中，所述网络通信端口 801 可以是与不同的通信协议进行绑定，从而可以发送或接收不同数据的虚拟端口。例如，所述网络通信端口可以是负责进行 web 数据通信的 80 号端口，也可以是负责进行 FTP 数据通信的 21 号端口，还可以是负责进行邮件数据通信的 25 号端口。此外，所述网络通信端口还可以是实体的通信接口或者通信芯片。例如，其可以为无线移动网络通信芯片，如 GSM、CDMA 等；其还可以为 Wifi 芯片；其还可以为蓝牙芯片。

15 [181] 在本实施方式中，所述处理器 802 可以按任何适当的方式实现。例如，处理器可以采取例如微处理器或处理器以及存储可由该（微）处理器执行的计算机可读程序代码（例如软件或固件）的计算机可读介质、逻辑门、开关、专用集成电路（Application Specific Integrated Circuit, ASIC）、可编程逻辑控制器和嵌入微控制器的形式等等。本说明书并不作限定。

20 [182] 在本实施方式中，所述存储器 803 可以包括多个层次，在数字系统中，只要能保存二进制数据的都可以是存储器；在集成电路中，一个没有实物形式的具有存储功能的电路也叫存储器，如 RAM、FIFO 等；在系统中，具有实物形式的存储设备也叫存储器，如内存条、TF 卡等。

25 [183] 本说明书实施例还提供一种用户端，包括处理器以及用于存储处理器可执行指令的存储器，所述处理器具体实施时可以根据指令执行以下步骤：显示数据输入界面；接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数，以及样本数据，其中，所述模型描述参数和所述样本数据用于确定基础模型，所述基础模型用于建立目标数据处理模型。

[184] 为了能够更加准确地完成上述指令，参阅图 9 所示，本说明书还提供了另一种

具体的用户端，其中，所述用户端包括显示器 901、处理器 902 以及存储器 903，上述结构通过内部线缆相连，以便各个结构可以进行具体的数据交互。

[185] 其中，所述显示器 901，具体可以用于显示数据输入界面。

5 [186] 所述处理器 902，具体可以用于接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数，以及样本数据，其中，所述模型描述参数和所述样本数据用于确定基础模型，所述基础模型用于建立目标数据处理模型。

[187] 所述存储器 903，具体可以用于存储相应的指令程序。

10 [188] 在本实施方式中，所述显示器 901 可以是将一定的电子文件通过特定的传输设备显示到屏幕上再反射到人眼的显示工具。所述显示器可以包括液晶 LCD 显示屏、阴极射线管 CRT 显示屏、发光二极管 LED 显示屏等。

[189] 本说明书实施例还提供了一种基于上述数据处理模型构建方法的计算机存储介质，所述计算机存储介质存储有计算机程序指令，在所述计算机程序指令被执行时实现：获取目标数据处理模型的模型描述参数和样本数据；根据所述模型描述参数和所述样本数据，确定基础模型；根据所述样本数据对所述基础模型进行训练，得到目标数据处理模型。

15

[190] 在本实施方式中，上述存储介质包括但不限于随机存取存储器（Random Access Memory, RAM）、只读存储器（Read-Only Memory, ROM）、缓存（Cache）、硬盘（Hard Disk Drive, HDD）或者存储卡（Memory Card）。所述存储器可以用于存储计算机程序指令。网络通信单元可以是依照通信协议规定的标准设置的，用于进行网络连接通信的接口。

20

[191] 在本实施方式中，该计算机存储介质存储的程序指令具体实现的功能和效果，可以与其它实施方式对照解释，在此不再赘述。

[192] 参阅图 10，在软件层面上，本说明书实施例还提供了一种数据处理模型构建装置，该装置具体可以包括以下的结构模块：

25 [193] 获取模块 1001，具体可以用于获取目标数据处理模型的模型描述参数和样本数据；

[194] 确定模块 1002，具体可以用于根据所述模型描述参数和所述样本数据，确定基础模型；

- [195] 训练模块 1003, 具体可以用于根据所述样本数据对所述基础模型进行训练, 得到目标数据处理模型。
- [196] 在一个实施方式中, 所述模型描述参数具体可以包括以下至少之一: 目标数据处理模型的名称、数据标签、数据特征、评估指标等。当然, 需要说明的是, 上述所列
- 5 举的模型描述参数只是一种示意性说明。具体实施时, 还可以引入其他类型的参数作为上述模型描述参数。对此, 本说明书不作限定。
- [197] 在一个实施方式中, 所述确定模块 1002, 具体可以包括以下结构单元:
- [198] 第一确定单元, 具体可以用于根据所述数据标签, 从预设的模型类型组中确定基础模型的模型类型, 其中, 所述预设的模型类型组中至少包括: 分类类型和回归预测
- 10 类型;
- [199] 选择单元, 具体可以用于从满足所述基础模型的模型类型的模型中, 选择出多个模型作为候选模型;
- [200] 建立单元, 具体可以用于根据所述样本数据和所述数据特征, 为多个候选模型中的各个候选模型建立预训练模型, 并确定各个预训练模型的模型参数;
- 15 [201] 第二确定单元, 具体可以用于根据所述评估指标, 确定各个预训练模型的适用度, 其中, 所述适用度用于表征基于预训练模型和该预训练模型对应的模型参数对所述样本数据进行数据处理的适用性程度;
- [202] 第三确定单元, 具体可以用于根据各个预训练模型的适用度, 从所述多个候选模型中确定出所述基础模型。
- 20 [203] 在一个实施方式中, 所述选择单元具体可以包括以下结构子单元:
- [204] 处理子单元, 具体可以用于分别利用预设的数据处理模型库中满足所述基础模型的模型类型的数据处理模型对所述样本数据进行数据处理, 得到数据处理模型的数据处理结果;
- [205] 统计子单元, 具体可以用于根据所述数据处理模型的数据处理结果, 统计数据
- 25 处理模型的评估指标;
- [206] 第一确定子单元, 具体可以用于将评估指标与目标数据处理模型的评估指标的差异程度小于预设程度阈值的数据处理模型确定为推荐模型;
- [207] 第二确定子单元, 具体可以用于将所述推荐模型所应用的基础模型确定为候选

模型。

[208] 在一个实施方式中，所述建立单元具体可以包括以下结构子单元：

[209] 提取子单元，具体可以用于从所述样本数据中提取预设比例的样本数据作为测试数据；

5 [210] 预处理子单元，具体可以用于对所述测试数据进行预处理，得到处理后的测试数据，其中，所述预处理包括：降维处理和/或清洗处理；

[211] 选择子单元，具体可以用于根据所述数据特征，从所述处理后的测试数据中选择出特征数据；

10 [212] 训练子单元，具体可以用于根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练，得到所述预训练模型。

[213] 在一个实施方式中，所述第三确定单元具体可以包括：第三确定子单元，具体可以用于将所述适用度最高的预训练模型所使用的模型确定为所述基础模型。

15 [214] 在一个实施方式中，所述第三确定单元具体还可以包括：第四确定子单元，具体可以用于将所述适用度大于预设适用度阈值的多个预训练模型所使用的模型进行组合，得到组合模型；并将所述组合模型确定为所述基础模型。

[215] 在一个实施方式中，所述模型描述参数还可以包括：基础模型的模型类型、样本数据的拆分规则等。这样可以为具备不同的专业能力的用户提供更多选择的模型描述参数，以便能够更加准确、快速地确定出基础模型和模型参数。

20 [216] 需要说明的是，上述实施例阐明的单元、装置或模块等，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本说明书时可以把各模块的功能在同一个或多个软件和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合
25 或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[217] 由上可见，本说明书实施例提供的数据处理模型构建装置，由于通过获取模块

有针对性地获取模型的描述参数，并基于模型的描述参数，通过确定模块准确地确定出用户的建模要求，并根据用户的建模要求自动匹配到合适的模型作为基础模型，再通过训练模块进行相应的数据处理模型的构建，从而使得用户不用再繁琐地去做测试、分析来确定合适的基础模型，降低了用户的操作难度，提高了处理效率。

5 [218] 虽然本说明书提供了如实施例或流程图所述的方法操作步骤，但基于常规或者无创造性的手段可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式，不代表唯一的执行顺序。在实际中的装置或客户端产品

10 其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、产品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、产品或者设备所固有的要素。在没有更多限制的情况下，并不排除在包括所述要素的过程、方法、产品或者设备中还存在另外的相同或等同要素。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

15 [219] 本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内部包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块

20 又可以是硬件部件内的结构。

[220] 本说明书可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构、类等等。也可以在分布式计算环境中实践本说明书，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算

25 环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[221] 通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本说明书可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本说明书的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如 ROM/RAM、磁碟、光盘等，包括若干指令用以

30 使得一台计算机设备（可以是个人计算机，移动终端，服务器，或者网络设备等等）执行

本说明书各个实施例或者实施例的某些部分所述的方法。

[222] 本说明书中的各个实施例采用递进的方式描述，各个实施例之间相同或相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。本说明书可用于众多通用或专用的计算机系统环境或配置中。例如：个人计算机、服务器计算机、
5 手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的电子设备、网络 PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[223] 虽然通过实施例描绘了本说明书，本领域普通技术人员知道，本说明书有许多变形和变化而不脱离本说明书的精神，希望所附的权利要求包括这些变形和变化而不脱
10 离本说明书的精神。

权利要求书

- 1.一种数据处理模型构建方法,包括:
 - 获取目标数据处理模型的模型描述参数和样本数据;
 - 根据所述模型描述参数和所述样本数据,确定基础模型;
- 5 根据所述样本数据对所述基础模型进行训练,得到目标数据处理模型。
 - 2.根据权利要求1所述的方法,所述模型描述参数包括以下至少之一:目标数据处理模型的名称、数据标签、数据特征、评估指标。
 - 3.根据权利要求2所述的方法,根据所述模型描述参数和所述样本数据,确定基础模型,包括:
 - 10 根据所述数据标签,从预设的模型类型组中确定基础模型的模型类型,其中,所述预设的模型类型组中至少包括:分类类型和回归预测类型;
 - 从满足所述基础模型的模型类型的模型中,选择出多个模型作为候选模型;
 - 根据所述样本数据和所述数据特征,为多个候选模型中的各个候选模型建立预训练模型,并确定各个预训练模型的模型参数;
 - 15 根据所述评估指标,确定各个预训练模型的适用度,其中,所述适用度用于表征基于预训练模型和该预训练模型对应的模型参数对所述样本数据进行数据处理的适用性程度;
 - 根据各个预训练模型的适用度,从所述多个候选模型中确定出所述基础模型。
 - 4.根据权利要求3所述的方法,从满足所述基础模型的模型类型的模型中,选择出
 - 20 多个模型作为候选模型,包括:
 - 利用预设的数据处理模型库中所应用的基础模型满足所述基础模型的模型类型的数据处理模型对所述样本数据进行数据处理,得到数据处理模型的数据处理结果;
 - 根据所述数据处理模型的数据处理结果,统计数据处理模型的评估指标;
 - 将评估指标与目标数据处理模型的评估指标的差异程度小于预设程度阈值的数据
 - 25 处理模型确定为推荐模型;
 - 将所述推荐模型所应用的基础模型确定为候选模型。
 - 5.根据权利要求3所述的方法,在根据所述数据标签,从预设的模型类型组中确定基础模型的模型类型后,所述方法还包括:
 - 确定所述样本数据的数据类型,其中,所述样本数据的数据类型包括以下至少之一:
 - 30 图片数据、文本数据;
 - 根据所述样本数据的数据类型,从所述基础模型的模型类型中筛选出与所述样本数

据的数据类型匹配的模型类型作为所述基础模型的模型类型。

6.根据权利要求3所述的方法,根据所述样本数据和所述数据特征,为多个候选模型中的各个候选模型建立预训练模型,包括:

从所述样本数据中提取预设比例的样本数据作为测试数据;

5 对所述测试数据进行预处理,得到处理后的测试数据,其中,所述预处理包括:降维处理和/或清洗处理;

根据所述数据特征,从所述处理后的测试数据中选择出特征数据;

根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练,得到所述预训练模型。

10 7.根据权利要求3所述的方法,根据所述评估指标,确定各个预训练模型的适用度,包括:

根据所述预训练模型和对应的模型参数,进行参数搜索和结构搜索,得到搜索结果;

根据所述搜索结果和所述评估指标,确定所述预训练模型的适用度。

15 8.根据权利要求7所述的方法,根据所述预训练模型和对应的模型参数,进行参数搜索和结构搜索,包括:

通过以下算法中的至少之一进行参数搜索和结构搜索:蒙特卡洛算法、网格算法、贝叶斯算法、遗传算法。

9.根据权利要求3所述的方法,根据各个预训练模型的适用度,从所述多个候选模型中确定出所述基础模型,包括:

20 将所述适用度最高的预训练模型所使用的模型确定为所述基础模型。

10.根据权利要求9所述的方法,根据各个预训练模型的适用度,从所述多个候选模型中确定出所述基础模型,还包括:

将所述适用度大于预设适用度阈值的多个预训练模型所使用的模型进行组合,得到组合模型;并将所述组合模型确定为所述基础模型。

25 11.根据权利要求2所述的方法,所述模型描述参数还包括:基础模型的模型类型、样本数据的拆分规则。

12.一种数据处理方法,包括:

获取待处理的目标数据;

30 调用目标数据处理模型对所述目标数据进行数据处理,其中,所述目标数据处理模型是根据所获取的目标数据处理模型的模型描述参数和样本数据,确定基础模型,再根据所述样本数据对所述基础模型进行训练得到的。

13.一种数据处理方法,包括:

显示数据输入界面;

接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数,以及样本数据,其中,所述模型描述参数和所述样本数据用于确定基础模型,所述基础模型用于

5 建立目标数据处理模型。

14.根据权利要求 13 所述的方法,所述模型描述参数包括以下至少之一:目标数据处理模型的名称、数据标签、数据特征、评估指标。

15.根据权利要求 13 所述的方法,在接收用户在所述数据输入界面输入的目标数据处理模型的模型描述参数,以及样本数据后,所述方法还包括:

10 响应展示操作,展示基于所述模型描述参数生成的执行文件;

接收对所述执行文件的修改操作,并根据所述修改操作更新所述执行文件。

16.一种数据处理模型构建装置,包括:

获取模块,用于获取目标数据处理模型的模型描述参数和样本数据;

确定模块,用于根据所述模型描述参数和所述样本数据,确定基础模型;

15 训练模块,用于根据所述样本数据对所述基础模型进行训练,得到目标数据处理模型。

17.根据权利要求 16 所述的装置,所述模型描述参数包括以下至少之一:目标数据处理模型的名称、数据标签、数据特征、评估指标。

18.根据权利要求 17 所述的装置,所述确定模块包括:

20 第一确定单元,用于根据所述数据标签,从预设的模型类型组中确定基础模型的模型类型,其中,所述预设的模型类型组中至少包括:分类类型和回归预测类型;

选择单元,用于从满足所述基础模型的模型类型的模型中,选择出多个模型作为候选模型;

25 建立单元,用于根据所述样本数据和所述数据特征,为多个候选模型中的各个候选模型建立预训练模型,并确定各个预训练模型的模型参数;

第二确定单元,用于根据所述评估指标,确定各个预训练模型的适用度,其中,所述适用度用于表征基于预训练模型和该预训练模型对应的模型参数对所述样本数据进行数据处理的适用性程度;

30 第三确定单元,用于根据各个预训练模型的适用度,从所述多个候选模型中确定出所述基础模型。

19.根据权利要求 18 所述的装置,所述选择单元包括:

处理子单元,用于分别利用预设的数据处理模型库中所应用的基础模型满足所述基础模型的模型类型的数据处理模型对所述样本数据进行数据处理,得到数据处理模型的数据处理结果;

统计子单元,用于根据所述数据处理模型的数据处理结果,统计数据处理模型的评估指标;

5 第一确定子单元,用于将评估指标与目标数据处理模型的评估指标的差异程度小于预设程度阈值的数据处理模型确定为推荐模型;

第二确定子单元,用于将所述推荐模型所应用的基础模型确定为候选模型。

20.根据权利要求 18 所述的装置,所述建立单元包括:

10 提取子单元,用于从所述样本数据中提取预设比例的样本数据作为测试数据;

预处理子单元,用于对所述测试数据进行预处理,得到处理后的测试数据,其中,所述预处理包括:降维处理和/或清洗处理;

选择子单元,用于根据所述数据特征,从所述处理后的测试数据中选择出特征数据;

15 训练子单元,用于根据所述特征数据分别对所述多个候选模型中的各个候选模型进行训练,得到所述预训练模型。

21.根据权利要求 18 所述的装置,所述第三确定单元包括:第三确定子单元,用于将所述适用度最高的预训练模型所使用的模型确定为所述基础模型。

22.根据权利要求 21 所述的装置,所述第三确定单元还包括:第四确定子单元,用于将所述适用度大于预设适用度阈值的多个预训练模型所使用的模型进行组合,得到组合模型;并将所述组合模型确定为所述基础模型。

23.根据权利要求 17 所述的装置,所述模型描述参数还包括:基础模型的模型类型、样本数据的拆分规则。

24.一种服务器,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器执行所述指令时实现权利要求 1 至 11 中任一项所述方法的步骤。

25 25.一种用户端,包括处理器以及用于存储处理器可执行指令的存储器,所述处理器执行所述指令时实现权利要求 13 至 15 中任一项所述方法的步骤。

26.一种计算机可读存储介质,其上存储有计算机指令,所述指令被执行时实现权利要求 1 至 11 中任一项所述方法的步骤。